

103	7	1.3	333	2	JC2329	translation initia
104	7	1.3	336	2	T50318	conserved hypot
105	7	1.3	339	2	F97190	phenylalanine-tRNA
106	7	1.3	341	1	G86315	hypothetical prote
107	7	1.3	345	1	KZEC	alkaline phosphat
108	7	1.3	345	2	G91079	alkaline phosphat
109	7	1.3	345	2	H85924	alkaline phosphat
110	7	1.3	346	2	AB3318	sodium-dependent p
111	7	1.3	351	2	E11372	probable translati
112	7	1.3	356	2	G84121	peptide chain rele
113	7	1.3	359	2	JC4297	MAPK-activated pro
114	7	1.3	363	2	H81410	hydrogenase isoenz
115	7	1.3	374	2	S46810	hypothetical prote
116	7	1.3	376	2	S63613	probable ATP-bind
117	7	1.3	378	2	T18486	hypothetical prote
118	7	1.3	381	2	C86358	Similar to zinc fi
119	7	1.3	383	1	S15624	B2 protein - human
120	7	1.3	387	2	A12569	hypothetical prote
121	7	1.3	388	2	D72713	hypothetical prote
122	7	1.3	389	2	H81806	O-succinylhomoseri
123	7	1.3	389	2	G81062	trans-sulfuration
124	7	1.3	390	2	C81852	L-lactate dehydrog
125	7	1.3	405	1	JH0795	calreticulin precu
126	7	1.3	420	2	AH2325	two-component sens
127	7	1.3	423	2	AD2455	hypothetical prote
128	7	1.3	424	2	H84906	hypothetical prote
129	7	1.3	424	2	S78610	hypothetical prote
130	7	1.3	425	2	T12473	hypothetical prote
131	7	1.3	432	1	FGLMGS	fibronogen gamma c
132	7	1.3	440	2	D95952	probable glycosylt
133	7	1.3	446	2	E70543	hypothetical prote
134	7	1.3	459	2	E75324	ArgE/DapE/AcyI fam
135	7	1.3	461	2	T14822	probable serine/th
136	7	1.3	468	2	D95411	probable decarboxy
137	7	1.3	474	2	T39567	serine/threonine-p
138	7	1.3	480	2	T08415	hypothetical prote
139	7	1.3	495	2	C95144	glucose-6-phosphat
140	7	1.3	495	2	A98012	glucose-6-phosphat
141	7	1.3	495	2	A44489	GT box-binding pro
142	7	1.3	496	2	C83122	probable aldehyde
143	7	1.3	506	2	C86164	probable aldehyde
144	7	1.3	507	2	G89908	Catalase imported
145	7	1.3	509	2	AH0578	citrate (pro-3S)-l
146	7	1.3	515	2	D71319	probable glucose-6
147	7	1.3	553	2	G71543	probable fructose-
148	7	1.3	557	2	JC6539	tumor necrosis fac
149	7	1.3	562	2	S75308	DNA ligase (EC 6.5
150	7	1.3	564	2	E87631	acyl-CoA dehydroge
151	7	1.3	565	2	A87634	ribosomal protein
152	7	1.3	570	2	D83964	fibronectin/fibrin
153	7	1.3	571	2	S52276	translation initia
154	7	1.3	572	2	G69877	fibronectin-bindin
155	7	1.3	578	2	S28432	hypothetical prote
156	7	1.3	578	2	E36832	Vi polysaccharide
157	7	1.3	578	2	AC1041	Vi polysaccharide
158	7	1.3	579	2	T15557	hypothetical prote
159	7	1.3	579	2	JH0830	160K Golgi antigen
160	7	1.3	584	2	E37146	indolepyruvate fer
161	7	1.3	585	2	E58403	H4 protein - human
162	7	1.3	590	2	S57594	hypothetical prote
163	7	1.3	601	1	A64222	heat shock protein
164	7	1.3	620	2	G71413	hypothetical 7K pr
165	7	1.3	630	2	S41314	hypothetical prote
166	7	1.3	648	2	T23864	hypothetical prote
167	7	1.3	675	2	S20819	collagen alpha 3(I)
168	7	1.3	687	2	JN0667	dnak-type molecula
169	7	1.3	703	2	AG0242	probable membrane
170	7	1.3	729	2	T50989	hypothetical prote
171	7	1.3	761	2	UC7821	beta-N-acetylhexos
172	7	1.3	764	2	T13302	myosin heavy chain
173	7	1.3	772	2	G02860	carnitine O-palmit
174	7	1.3	772	2	S65532	carnitine palmitoy
175	7	1.3	776	2	T19900	hypothetical prote

[illegible]

caldesmon - human
glutamine-tRNA ligase
hypothetical protein
conserved hypothetical
hypothetical protein
hypothetical protein
conserved hypothetical
probable adenylate
env polyprotein pr
hypothetical protein
chloroplast outer
probable membrane
kinase-like protein
the two components
the two components
p50B/p97 (Lyt-10)
exocyst complex pr
isoletucine-tRNA li
hypothetical protein
hypothetical protein
hypothetical protein
probable Na⁺/K⁺/Cl⁻
hypothetical protein
mucin 5AC (clone I
translation initia
ribonuclease B PA2
carbamoyl-phosphat
translation initia
hypothetical colle
sterol regulatory
sterol regulatory
DNA topoisomerase
DNA topoisomerase
hypothetical protein
hypothetical protein
hypothetical protein
glutamate rich pro
two-component hyb
probable metal bin
dna-directed RNA bi
hypothetical protei
alkaline phosphatase
hypothetical protei
probable metal bin
chloroplast outer
hypothetical prote
probable DNA (cyc
Ras GTPase activat
embryonic muscle m
slow myosin heavy
myosin alpha heav
myosin heavy chain
myosin heavy chain
hypothetical prote
hypothetical prote
hypothetical prote
hypothetical prote
seven-pass transme
CREB-binding prote
probable non-ribos
dynein beta heavy
gamma-2-gliadin p25
G-alpha-1 protein
hypothetical prote
hypothetical prote
storage protein -
light-harvesting p
hypothetical prote
hypothetical prote
ydaS protein - Esc
probable bacterio
hypothetical prote
hypothetical prote
proline/betaine tr
hypothetical prote
hypothetical prote
hypothetical 7K pr

249 6 1.1 69 2 G82028 probable integral
250 6 1.1 71 2 E83654 hypothetical prote
251 6 1.1 72 2 T25597 hypothetical prote
252 6 1.1 76 2 D83182 protein W6E11.5
253 6 1.1 77 2 H70642 probable ribosomal
254 6 1.1 78 2 F72295 hypothetical prote
255 6 1.1 78 2 A13213 conserved hypothet
256 6 1.1 79 2 T02461 hypothetical prote
257 6 1.1 79 2 AG0340 hypothetical prote
258 6 1.1 79 2 AD1370 hypothetical prote
259 6 1.1 79 2 B81323 hypothetical prote
260 6 1.1 80 2 JQ0664 exodeoxyribonuclea
261 6 1.1 80 2 D90688 exonuclease VII sm
262 6 1.1 80 2 AH0554 exonuclease VII, s
263 6 1.1 80 2 H85538 hypothetical prote
264 6 1.1 81 2 H83341 conserved hypothet
265 6 1.1 81 2 C82398 conserved hypothet
266 6 1.1 82 2 C84086 hypothetical prote
267 6 1.1 82 2 S11843 virE1 protein - Ag
268 6 1.1 83 2 AB3251 virA/G regulated p
269 6 1.1 84 2 S35758 ribosomal protein
270 6 1.1 84 2 A48045 ribosomal protein
271 6 1.1 85 2 A97934 degenerate transpo
272 6 1.1 86 2 G90675 probable ATP bindi
273 6 1.1 86 2 B85526 probable ATP bindi
274 6 1.1 89 2 T51589 probable flavonol
275 6 1.1 92 2 S17443 ribosomal protein
276 6 1.1 92 2 B95032 conserved hypothet
277 6 1.1 92 2 E97903 conserved hypothet
278 6 1.1 93 2 I45876 collagen alpha 1(I
279 6 1.1 93 2 G64641 conserved hypothet
280 6 1.1 93 2 H71872 glu-tRNA amidotran
281 6 1.1 94 2 G86658 hypothetical prote
282 6 1.1 95 2 T03068 hypothetical prote
283 6 1.1 95 2 T13201 hypothetical prote
284 6 1.1 96 2 A31895 nonhistone chromos
285 6 1.1 96 2 G70117 conserved hypothet
286 6 1.1 96 2 F64319 hypothetical prote
287 6 1.1 96 2 T23262 hypothetical prote
288 6 1.1 97 2 C69321 conserved hypothet
289 6 1.1 98 2 B64763 probable alpha hel
290 6 1.1 98 2 D90688 probable alpha hel
291 6 1.1 98 2 H85530 probable alpha hel
292 6 1.1 99 2 T36223 hypothetical prote
293 6 1.1 99 2 G82935 conserved hypothet
294 6 1.1 100 2 F81677 conserved hypothet
295 6 1.1 100 2 AB1035 probable phage tai
296 6 1.1 100 2 AG0836 probable phage tai
297 6 1.1 100 2 AI0829 probable phage tai
298 6 1.1 102 2 F72644 hypothetical prote
299 6 1.1 103 2 I37454 gene HB2 protein -
300 6 1.1 104 2 AC0162 probable membrane

ALIGNMENTS

RESULT 1
JC7222
77K muscle-derived protein - chicken
C:Species: Gallus gallus (chicken)
C>Date: 09-Jun-2000 #sequence_revision 09-Jun-2000 #text_change 21-Jul-2000
C:Accession: JC7222
R:Oyeda, A.; Fukui, I.; Fujimori, K.; Kiyosue, K.; Nishimune, H.; Kasai, M.; Taguchi, T.
Biochem. Biophys. Res. Commun. 269, 564-569, 2000
A:Title: MDP77: A novel neurite-outgrowth-promoting protein predominantly expressed in c
A:Reference number: JC7222; MUID:20175243; PMID:10708594
A:Accession: JC7222
A:Molecule type: mRNA
A:Residues: 1-676 <UVE>
A:Cross-references: GB:D89999; NID:G7619883; PIDN:BA94755.1; PID:G7619884
A:Experimental source: crus muscle
C:Comment: This protein, a glycoprotein and a neurite-outgrowth-promoting protein, is im

C:Keywords: coiled coil; glycoprotein; leucine zipper; muscle
Query Match 3.2%; Score 17; DB 2; Length 676;
Best Local Similarity 100.0%; Pred. No. 2e-08;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 220 LARSKLESICRELQHN 236
|||||
DB 207 LARSKLESICRELQHN 223
|||||
RESULT 2
S50611
hypothetical protein YER108c - yeast (Saccharomyces cerevisiae)
C:Species: Saccharomyces cerevisiae
C>Date: 28-May-1993 #sequence_revision 24-Feb-1995 #text_change 23-Mar-2001
C:Accession: S50611
R:Dieterich, F.S.
A:Description: The sequence of S. cerevisiae cosmids 9781, 8198, 9115, 9981, and lambda da
A:Reference number: S50437
A:Accession: S50611
A:Molecule type: DNA
A:Residues: 1-594 <DIR>
A:Cross-references: EMBL:U18916; NID:gl394128; PIDN:AAC03206.1; PID:G603347; MIPS:YER10
C:Genetics:
A:Gene: SGD:YLO8
A:Cross-references: SGD:S0000910; MIPS:YER108c
A:Map position: 5R
Query Match 1.7%; Score 9; DB 2; Length 594;
Best Local Similarity 100.0%; Pred. No. 2.2;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 164 NTLSTPEEK 172
|||||
DB 466 NTLSTPEEK 474
|||||
RESULT 3
I46707
translation initiation factor eIF4-gamma - rabbit
C:Species: Oryctolagus cuniculus (domestic rabbit)
C>Date: 14-Feb-1997 #sequence_revision 14-Feb-1997 #text_change 31-Mar-2000
C:Accession: I46707
R:Yan, R.; He, W.; Rhoads, R.E.
J. Biol. Chem. 268, 19200-19203, 1993
A:Title: Mapping the cleavage site in protein synthesis initiation factor eIF 4 gamma o
A:Reference number: I46707; MUID:93374895; PMID:8396129
A:Accession: I46707
A>Status: preliminary; translated from GB/EMBL/DDBJ
A:Molecule type: mRNA
A:Residues: 1-1402 <YAN>
A:Cross-references: GB:I22090; NID:G404774; PIDN:AAA31242.1; PID:G404775
Query Match 1.7%; Score 9; DB 2; Length 1402;
Best Local Similarity 100.0%; Pred. No. 5;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 56 SEELSPQLE 64
|||||
DB 1243 SEELSPQLE 1251
|||||
RESULT 4
T29728
hypothetical protein K03B4.7 - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C>Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 15-Oct-1999
C:Accession: T29728
R:Du, Z.; Le, T.T.; Kemp, K.
submitted to the EMBL Data Library, April 1996
A:Description: The sequence of C. elegans cosmid K03B4.

A:Reference number: Z20673
A:Accession: T29728
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-162 <DUZ>
A:Cross-references: EMBL:U55370; PIDN:AAA97999.1; GSPDB:GN00023; CESP:K03B4.7
A:Experimental source: strain Bristol N2; clone K03B4
C:Genetics:
A:Gene: CESP:K03B4.7
A:Map position: 5
A:Introns: 58/3; 105/3

Query Match 1.5%; Score 8; DB 2; Length 162;
Best Local Similarity 100.0%; Pred. No. 6.8;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 290 AERLUKLI 297
|||
Db 89 AERLUKLI 96

RESULT 5
T32354
hypochemical protein C08E3.4 - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C:Date: 29-Oct-1999 #sequence_revision 29-Oct-1999 #text_change 29-Oct-1999
C:Accession: T32354
R:Miller, N.; Kramer, J.; Keppler, D.
A:Description: The sequence of C. elegans cosmid C08E3.
A:Reference number: Z21155
A:Accession: T32354
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-211 <ML>
A:Cross-references: EMBL:AF025457; PIDN:AA070965.1; GSPDB:GN00020; CESP:C08E3.4
A:Experimental source: strain Bristol N2; clone C08E3
C:Genetics:
A:Gene: CESP:C08E3.4
A:Map position: 2
A:Introns: 36/3; 149/3

Query Match 1.5%; Score 8; DB 2; Length 211;
Best Local Similarity 100.0%; Pred. No. 8.7;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 290 AERLUKLI 297
|||
Db 81 AERLUKLI 88

RESULT 6
T52558
translation elongation factor eEF1alpha (clone 2) [validated] - Arabidopsis thaliana
C:Species: Arabidopsis thaliana (mouse-ear cress)
C:Date: 24-Oct-2000 #sequence_revision 24-Oct-2000 #text_change 07-Jul-2003
C:Accession: T52558
R:Hericourt, F.; Jupin, I.
A:Title: Molecular cloning and characterization of the Arabidopsis thaliana alpha-subunit
A:Reference number: Z26114
A:Accession: T52558
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 1-224 <HER>
A:Cross-references: EMBL:AJ249597; PIDN:CAB64730.1
C:Genetics:
A:Gene: eEF1alpha2
A:Function:
A:Description: involved in translation elongation; able to complement a mutant yeast strain
C:Superfamily: translation elongation factor eEF-1 beta chain; translation elongation factor

Query Match 1.5%; Score 8; DB 2; Length 224;

Best Local Similarity 100.0%; Pred. No. 9.2;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 21 AAPAVEAE 28
|||
Db 84 AAPAVEAE 91

RESULT 7
E87236
probable membrane protein. [imported] - Mycobacterium leprae
C:Species: Mycobacterium leprae
C:Date: 20-Apr-2001 #sequence_revision 20-Apr-2001 #text_change 10-May-2001
C:Accession: E87236
R:Coile, S.T.; Egimieier, K.; Parkhill, J.; James, K.D.; Thomson, N.R.; Wheeler, P.R.; F
R.; Davies, R.M.; Devlin, K.; Duthoy, S.; Feltwell, T.; Fraser, A.; Hamlin, E.; Holroyd
eam, M.A.; Rutherford, K.M.
A:Title: Massive gene decay in the leprosy bacillus.
A:Reference number: A86909; MUID:21128732; PMID:11234002
A:Accession: E87236
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-229 <STO>
A:Cross-references: GB:AL450380; NID:gl3093841; PIDN:CAC32147.1; GSPDB:GN00147
C:Genetics:
A:Gene: ML2615
C:Superfamily: Mycobacterium tuberculosis hypothetical protein Rv0200

Query Match 1.5%; Score 8; DB 2; Length 229;
Best Local Similarity 100.0%; Pred. No. 9.4;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 18 PSQAAPAV 25
|||
Db 200 PSQAAPAV 207

RESULT 8
S71315
deoxyguanosine kinase (EC 2.7.1.113) precursor - human
C:Species: Homo sapiens (man)
C:Date: 12-Feb-1998 #sequence_revision 13-Mar-1998 #text_change 20-Jun-2000
C:Accession: S71315; S78432
R:Wang, L.; Hellman, U.; Eriksson, S.
FEBS Lett. 390, 39-43, 1996
A:Title: Cloning and expression of human mitochondrial deoxyguanosine kinase cDNA.
A:Reference number: S71315; MUID:96314545; PMID:8706825
A:Accession: S71315
A:Molecule type: mRNA
A:Residues: 1-260 <WAW>
A:Cross-references: EMBL:X97386; PIDN:CAA66054.1
A:Experimental source: tissue brain
R:Wang, L.; Hellman, U.; Eriksson, S.
submitted to the EMBL Data Library, April 1996
A:Description: Cloning and expression of human deoxyguanosine kinase cDNA.
A:Reference number: S78432
A:Accession: S78432
A:Molecule type: mRNA
A:Residues: 1-18, 'R', 20-260 <WAW>
A:Cross-references: EMBL:X97386; PIDN:CAA66054.1
A:Experimental source: brain
C:Genetics:
A:Gene: dgk
A:Genome: nuclear
C:Superfamily: human deoxycytidine kinase
C:Keywords: mitochondrion; phosphotransferase
P:1-22/Domain: transit peptide (mitochondrion) #status predicted <TMP>
E:23-260/Product: transit peptide (mitochondrion)
P:125-132/Region: DRS motif
P:185-191/Region: arginine-rich

Query Match 1.5%; Score 8; DB 2; Length 260;
 Best Local Similarity 100.0%; Pred. No. 11;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 245 QRAPEE 252
 |||||
 Db 188 QRAPEE 195
 |||||

RESULT 9
 C88638
 protein F58F6.1 [imported] - Caenorhabditis elegans
 C;Species: Caenorhabditis elegans
 C;Date: 10-May-2001 #sequence_revision 10-May-2001 #text_change 15-Sep-2003
 C;Accession: C88638
 R;Anonymous, The C. elegans Sequencing Consortium.
 Science 283, 2012-2018, 1998
 A;Title: Genome sequence of the nematode C. elegans: a platform for investigating biology
 A;Reference number: A75000; MUID:99069613; PMID:9851916
 A;Note: see websites genome.wustl.edu/gsc/C_elegans/ and www.sanger.ac.uk/Projects/C_ele
 A;Note: published errata appeared in Science 283, 35, 1999; Science 283, 2103, 1999; and
 A;Accession: C88638
 A;Status: preliminary
 A;Molecule type: DNA
 A;Residues: 1-281 <STO>
 A;Cross-references: GB:chr_IV; PIDN:AAB88358.1; PID:g2662599; GSPDB:GN00022; CESP:F58F6.
 C;Genetics:
 A;Gene: F58F6.1
 A;Map position: 4

Query Match 1.5%; Score 8; DB 2; Length 281;
 Best Local Similarity 100.0%; Pred. No. 11;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 77 GGPEDGA 84
 |||||
 Db 241 GGPEDGA 248
 |||||

RESULT 10
 F03985
 hypothetical protein F6G3.120 - Arabidopsis thaliana
 C;Species: Arabidopsis thaliana (mouse-ear cress)
 C;Date: 11-Jun-1999 #sequence_revision 11-Jun-1999 #text_change 22-Oct-1999
 R;Accession: T08985
 R;Bavan, M.; Murphy, G.; Ridley, P.; Hudson, S.; Bancroft, I.; Mewes, H.W.; Mayer, K.F.X.
 submitted to the Protein Sequence Database, May 1999
 A;Reference number: Z16520
 A;Accession: F03985
 A;Molecule type: DNA
 A;Residues: 1-312 <BEV>
 A;Cross-references: EMBL:AL078464; GSPDB:GN00062; ATSP:F6G3.120
 A;Experimental source: cultivar Columbia; BAC clone F6G3
 C;Genetics:
 A;Gene: ATSP:F6G3.120
 A;Map position: 4
 A;Introns: 55/1; 94/3; 115/3; 151/3; 168/3; 189/3; 241/3; 264/3; 285/2

Query Match 1.5%; Score 8; DB 2; Length 312;
 Best Local Similarity 100.0%; Pred. No. 13;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 326 QQAQEMLK 333
 |||||
 Db 240 QQAQEMLK 247
 |||||

RESULT 11
 E71909
 hypothetical protein jhp0612 - Helicobacter pylori (strain J99)
 C;Species: Helicobacter pylori
 A;Variety: strain J99
 C;Date: 12-Feb-1999 #sequence_revision 12-Feb-1999 #text_change 08-Oct-1999

C;Accession: E71909
 R;Alm, R.A.; Ling, L.S.L.; Moir, D.T.; King, B.L.; Brown, E.D.; Doig, P.C.; Smith, D.R.;
 Ives, C.; Gibson, R.; Merberg, D.; Mills, S.D.; Jiang, Q.; Taylor, D.E.; Vovis, G.F.;
 Nature 397, 176-180, 1999
 A;Title: Genomic sequence comparison of two unrelated isolates of the human gastric pat
 A;Reference number: A71800; MUID:99120557; PMID:9923682
 C;Accession: E71909
 A;Status: preliminary
 A;Molecule type: DNA
 A;Residues: 1-450 <ARN>
 A;Cross-references: GB:AE001493; GB:AE001439; NID:g4155161; PIDN:AAD06193.1; PID:g41551
 A;Experimental source: strain J99
 C;Genetics:
 A;Gene: jhp0612

Query Match 1.5%; Score 8; DB 2; Length 450;
 Best Local Similarity 100.0%; Pred. No. 18;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 63 LBDILSTY 70
 |||||
 Db 283 LBDILSTY 290
 |||||

RESULT 12
 D88710
 protein C43G2.1 [imported] - Caenorhabditis elegans
 C;Species: Caenorhabditis elegans
 C;Date: 10-May-2001 #sequence_revision 10-May-2001 #text_change 10-May-2001
 C;Accession: D88710
 R;Anonymous, The C. elegans Sequencing Consortium.
 Science 282, 2012-2018, 1998
 A;Title: Genome sequence of the nematode C. elegans: a platform for investigating biolo
 A;Reference number: A75000; MUID:99069613; PMID:9851916
 A;Note: see websites genome.wustl.edu/gsc/C_elegans/ and www.sanger.ac.uk/Projects/C_el
 A;Note: published errata appeared in Science 283, 35, 1999; Science 283, 2103 1999; an
 A;Accession: D88710
 A;Status: preliminary
 A;Molecule type: DNA
 A;Residues: 1-467 <STO>
 A;Cross-references: GB:chr_IV; PIDN:AAB09107.1; PID:g1572756; GSPDB:GN00022; CESP:C43G2
 C;Genetics:
 A;Gene: C43G2.1
 A;Map position: 4

Query Match 1.5%; Score 8; DB 2; Length 467;
 Best Local Similarity 100.0%; Pred. No. 18;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 173 LAALCKXY 180
 |||||
 Db 22 LAALCKXY 29
 |||||

RESULT 13
 H97769
 hypothetical protein murd [imported] - Rickettsia conorii (strain Malish 7)
 C;Species: Rickettsia conorii
 C;Date: 30-Sep-2001 #sequence_revision 30-Sep-2001 #text_change 22-Oct-2001
 C;Accession: H97769
 R;Ogata, H.; Audic, S.; Renesto-Audiffren, P.; Pournier, P.E.; Barbe, V.; Samson, D.; R
 Science 293, 2093-2098, 2001
 A;Title: Mechanisms of Evolution in Rickettsia conorii and Rickettsia prowazekii.
 A;Reference number: A97700; MUID:2142074; PMID:11557993
 C;Accession: H97769
 A;Status: preliminary
 A;Molecule type: DNA
 A;Residues: 1-500 <KUR>
 A;Cross-references: GB:AB006914; PIDN:AAL03098.1; PID:g15619640; GSPDB:GN00173
 C;Genetics:
 A;Gene: murD
 C;Superfamily: UDP-N-acetylmuramate-alanine ligase

Query Match 1.5%; Score 8; DB 2; Length 500;
Best Local Similarity 100.0%; Pred. No. 19;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 85 QGEPAPPE 92
Db 278 QGEPAPPE 285
|||||

RESULT 14
H75514
glutamyl-tRNA synthetase - Deinococcus radiodurans (strain R1)
C:Species: Deinococcus radiodurans
C>Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 17-Mar-2000
C:Accession: H75514
X:White, O.; Eisen, J.A.; Heidelberg, J.P.; Hickey, E.K.; Peterson, J.D.; Dodson, R.J.;
M.; Shen, M.; Vamatrav, J.G.; Lam, P.; McDonald, L.; Utterback, T.; Zalewski, C.; Ma
S.; Smith, H.O.; Venter, J.C.; Fraser, C.M.
Science 286, 1571-1577, 1999
A>Title: Genome sequence of the radioresistant bacterium Deinococcus radiodurans R1.
A:Reference number: A75250; MUID:20036896; PMID:10567266
A:Accession: H75514
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-525 <WHI>
A:Cross-references: GB:AB001907; GB:AE000513; NID:96458162; PIDN:AAFI0063.1; PID:9645817
A:Experimental source: strain R1
C:Genetics:
A:Gene: DR0485
A:Map position: 1
C:Superfamily: glutamate-tRNA ligase; glutamine-tRNA ligase homology

Query Match 1.5%; Score 8; DB 2; Length 525;
Best Local Similarity 100.0%; Pred. No. 20;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 496 APSSPRVT 503
Db 45 APSSPRVT 52
|||||

RESULT 15
A55887
caldesmon, non-muscle - rat
C:Species: Rattus norvegicus (Norway rat)
C>Date: 23-Mar-1995 #sequence_revision 23-Mar-1995 #text_change 13-Aug-1999
C:Accession: A55887
R:Yamashiro, S.; Yamakita, Y.; Yoshida, K.; Takiguchi, K.; Matsumura, F.
J. Biol. Chem. 270, 4023-4030, 1995
A>Title: Characterization of the COOH terminus of non-muscle caldesmon mutants lacking
A:Reference number: A55887; MUID:95181370; PMID:7876150
A:Accession: A55887
A>Status: preliminary; nucleic acid sequence not shown
A:Molecule type: mRNA
A:Residues: 1-531 <YAM>
A:Cross-references: GB:U18419; NID:G622966; PIDN:AAA68521.1; PID:G622967
A>Note: authors translated the codon GCC for residue 68 as Val
C:Superfamily: caldesmon
C:Keywords: phosphoprotein

Query Match 1.5%; Score 8; DB 2; Length 531;
Best Local Similarity 100.0%; Pred. No. 21;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 247 AREEEKR 254
Db 334 AREEEKR 341
|||||

RESULT 16
C36811
hypothetical protein ORF50 - saimiriine herpesvirus 1 (strain 11)
C:Species: saimiriine herpesvirus 1

A>Note: host Saimiri sciureus (common squirrel monkey)
C>Date: 16-Oct-1992 #sequence_revision 16-Oct-1992 #text_change 27-Jan-1995
C:Accession: C36811
R:Albrecht, J.
submitted to the EMBL Data Library, January 1992
A>Description: Primary structure of the herpesvirus saimiri genome.
A:Reference number: A36806
A:Accession: C36811
A:Molecule type: DNA
A:Residues: 1-535 <ALB>
A:Cross-references: GB:X64346
J:Albrecht, J.C.; Nicholas, J.; Biller, D.; Cameron, K.R.; Biesinger, B.; Newman, C.; W
J. Virol. 66, 5047-5058, 1992
A>Title: Primary structure of the herpesvirus saimiri genome.
A:Reference number: A37309; MUID:92333688; PMID:1321287
A:Contents: annotation; protein-coding frames
A>Note: neither protein nor nucleotide sequence is given
C:Genetics:
A:Gene: 50

Query Match 1.5%; Score 8; DB 2; Length 535;
Best Local Similarity 100.0%; Pred. No. 21;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 151 GLGKEITL 158
Db 187 GLGKEITL 194
|||||

RESULT 17
A37994
RFl protein - saimiriine herpesvirus 1
C:Species: saimiriine herpesvirus 1
C>Date: 23-Aug-1991 #sequence_revision 30-Jan-1993 #text_change 04-Mar-1994
C:Accession: A37994
R:Nicholas, J.; Coles, L.S.; Newman, C.; Honess, R.W.
J. Virol. 65, 2457-2466, 1991
A>Title: Regulation of the herpesvirus saimiri (HVS) delayed-early 110-kilodalton prom
A:Reference number: A37994; MUID:91202588; PMID:1850023
A:Accession: A37994
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-535 <NIC>
A:Cross-references: GB:M60850

Query Match 1.5%; Score 8; DB 2; Length 535;
Best Local Similarity 100.0%; Pred. No. 21;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 151 GLGKEITL 158
Db 187 GLGKEITL 194
|||||

RESULT 18
E90364
hypothetical protein SSO1981 [imported] - Sulfolobus solfataricus
C:Species: Sulfolobus solfataricus
C>Date: 24-May-2001 #sequence_revision 24-May-2001 #text_change 24-May-2001
C:Accession: E90364
R:She, Q.; Singh, R.K.; Confalonieri, F.; Zivanovic, Y.; Allard, G.; Awayez, M.J.; Chan
Jong, I.; Jeffries, A.C.; Kozera, C.J.; Medina, N.; Peng, X.; Thi-Ngoc, H.P.; Redder,
arrett, R.A.; Ragan, M.A.; Sengen, C.W.; Van der Oost, J.
submitted to GenBank, April 2001
A>Description: Sulfolobus solfataricus complete genome.
A:Reference number: A99139
A:Accession: E90364
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-568 <GR>
A:Cross-references: GB:AE006641; NID:g13815258; PIDN:AAK42172.1; GSPDB:GN0015;
C:Genetics:
A:Gene: SSO1981

Query Match 1.5%; Score 8; DB 2; Length 568;
 Best Local Similarity 100.0%; Pred. No. 22;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 404 KKIKKLEK 411
 |||||
 Db 91 KKIKKLEK 98
 |||||

RESULT 19
 S62141
 N:Alternative names: protein IIB chain B'' - yeast (Saccharomyces cerevisiae)
 N:Alternative names: protein N2682; protein YNL039W
 C:Species: Saccharomyces cerevisiae
 C:Date: 10-Apr-1996 #sequence_revision 19-Apr-1996 #text_change 21-Jul-2000
 C:Accession: S62141; S62961; S62912
 R:Kasavatis, G.A.; Nguyen, S.T.; Kobayashi, R.; Kumar, A.; Geiduschek, E.P.; Pisano, M.
 Proc. Natl. Acad. Sci. U.S.A. 92, 9786-9790, 1995
 A:Title: Cloning, expression, and function of TFC5, the gene encoding the B'' component
 A:Reference number: S62141; PMID:7568218
 A:Accession: S62141
 A:Status: nucleic acid sequence not shown
 A:Molecule type: DNA
 A:Residues: 1-594 <KAS>
 A:Cross-references: EMBL:U31819; NID:G1019114; PIDN:AAC49073.1; PID:G1019115
 R:Dueterhoef, A.; Floeth, M.; Fritz, C.; Heuss-Neitzel, D.; Hilbert, H.; Moestl, D.
 submitted to the Protein Sequence Database, April 1996
 A:Reference number: S62944
 A:Accession: S62961
 A:Molecule type: DNA
 A:Cross-references: EMBL:271315; NID:G1301889; PID:e239882; PID:G1301890; MIPS:YNL039W
 A:Experimental source: strain S288C
 R:Rueh, J.; Conesa, C.; Dieci, G.; Lefebvre, O.; Dueterhoef, A.; Ottonello, S.; Sente
 EMBO J. 15, 1941-1949, 1996
 A:Title: A suppressor of mutations in the class III transcription system encodes a compo
 A:Reference number: S6212; PMID:96203116; PMID:8617241
 A:Accession: S62912
 A:Status: nucleic acid sequence not shown
 A:Molecule type: DNA
 A:Residues: 1-594 <RUE>
 A:Cross-references: EMBL:U38415; NID:G1145357; PIDN:AAC49364.1; PID:G1145358
 A:Experimental source: strain S288C
 C:Keywords: transcription initiation

Query Match 1.5%; Score 8; DB 2; Length 594;
 Best Local Similarity 100.0%; Pred. No. 23;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 249 EEEERKE 256
 |||||
 Db 334 EEEERKE 341
 |||||

RESULT 20
 S65169
 N:Alternative names: protein YPL158c - yeast (Saccharomyces cerevisiae)
 N:Alternative names: hypothetical protein P2570
 C:Species: Saccharomyces cerevisiae
 C:Date: 10-Dec-1994 #sequence_revision 31-May-1996 #text_change 19-Apr-2002
 C:Accession: S65169; S69441
 R:Purnelle, B.; Coster, F.; Goffeau, A.
 submitted to the Protein Sequence Database, May 1996
 A:Reference number: S65154
 A:Accession: S65169
 A:Molecule type: DNA
 A:Residues: 1-758 <PUR>
 A:Cross-references: EMBL:273514; NID:G1370335; PID:e247047; PID:G1370336; MIPS:YPL158C
 A:Experimental source: strain S288C (AB972)

R:Purnelle, B.; Comblez, S.; Coster, F.; Naveau, F.; Goffeau, A.
 submitted to the EMBL Data Library, March 1996
 A:Description: The sequence of 55 kb on the left arm of yeast chromosome XVI identifies
 a gene to the human phosphotyrosyl phosphatase activator PTPA and a homologue to the plan
 A:Reference number: S69428
 A:Accession: S69441
 A:Molecule type: DNA
 A:Residues: 1-758 <PUW>
 A:Cross-references: EMBL:X96770; NID:G1403537; PID:e239043; PID:G1403551
 C:Genetics:
 A:Cross-references: SGD:S0006079
 A:Map position: 16L

Query Match 1.5%; Score 8; DB 2; Length 758;
 Best Local Similarity 100.0%; Pred. No. 29;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 249 EEEERKE 256
 |||||
 Db 687 EEEERKE 694
 |||||

RESULT 21
 T32183
 N:Alternative names: protein K09C6.1 - Caenorhabditis elegans
 C:Species: Caenorhabditis elegans
 C:Date: 29-Oct-1999 #sequence_revision 29-Oct-1999 #text_change 29-Oct-1999
 C:Accession: T32183
 R:Goela, D.; Harper, M.
 submitted to the EMBL Data Library, September 1997
 A:Description: The sequence of C. elegans cosmid K09C6.
 A:Reference number: Z21131
 A:Accession: T32183
 A:Status: preliminary; translated from GB/EMBL/DBJ
 A:Molecule type: DNA
 A:Residues: 1-761 <GOE>
 A:Cross-references: EMBL:AF022975; PIDN:AAB70668.1; GSPDB:GN00023; CESP:K09C6.1
 A:Experimental source: strain Bristol N2; clone K09C6
 C:Genetics:
 A:Gene: CESP:K09C6.1
 A:Map position: 5
 A:Introns: 48/3; 91/3; 123/2; 146/2; 200/2; 238/3; 269/1; 302/1; 481/3; 591/1; 639/1

Query Match 1.5%; Score 8; DB 2; Length 761;
 Best Local Similarity 100.0%; Pred. No. 29;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 181 AELLEER 188
 |||||
 Db 710 AELLEER 717
 |||||

RESULT 22
 T52292
 N:Alternative names: Clp (EC 3.4.21.92) ATP-binding chain C, chloroplast [imported] - Arabidop
 N:Alternative names: ATP-dependent clp proteinase chain C
 C:Species: Arabidopsis thaliana (mouse-ear cress)
 C:Date: 20-Oct-2000 #sequence_revision 20-Oct-2000 #text_change 01-Dec-2000
 C:Accession: T52292
 R:Clarke, A.K.

submitted to the EMBL Data Library, September 1997
 A:Description: A cDNA clone coding for the chloroplast ClpC protein from Arabidopsis th
 A:Reference number: Z26020
 A:Accession: T52292
 A:Status: preliminary; translated from GB/EMBL/DBJ
 A:Molecule type: mRNA
 A:Residues: 1-928 <CUA>
 A:Cross-references: EMBL:AF022909; PIDN:AAC04687.1
 C:Superfamily: endopeptidase Clp ATP-binding chain
 C:Keywords: ATP; chloroplast; hydrolase; nucleotide binding; serine proteinase

Query Match 1.5%; Score 8; DB 2; Length 928;
 Best Local Similarity 100.0%; Pred. No. 35;

Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 127 EEIRQSDS 134
| | | | |
Db 359 EEIRQSDS 366

RESULT 23

T52456
endopeptidase Clp ATP-binding chain C [imported] - Arabidopsis thaliana
C:Species: Arabidopsis thaliana (mouse-ear cress)
C>Date: 24-Oct-2000 #sequence_revision 24-Oct-2000 #text_change 08-Dec-2000
C:Accession: T52456
R:Nakabayashi, K.; Ito, M.; Kiyosue, T.; Shinozaki, K.; Watanabe, A.
Plant Cell Physiol. 40, 504-514, 1999
A:Title: Identification of clp genes expressed in senescing Arabidopsis leaves.
A:Reference number: Z2128; MUID:93356780; PMID:10427773
A:Accession: T52456
A:Status: preliminary;
A:Molecule type: mRNA
A:Residues: 1-952 <NAK>
A:Cross-references: EMBL:AB022324; NID:G5360573; PIDN:BA82062.1; PID:G5360574
A:Experimental source: strain Columbia
C:Genetics:
A:Gene: clpC
C:Superfamily: endopeptidase Clp ATP-binding chain

Query Match 1.5%; Score 8; DB 2; Length 952;
Best Local Similarity 100.0%; Pred. No. 36;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 127 EEIRQSDS 134
| | | | |
Db 381 EEIRQSDS 388

RESULT 24

T49283
AtClpC - Arabidopsis thaliana
N/Alternate names: protein T21J18.140
C:Species: Arabidopsis thaliana (mouse-ear cress)
C>Date: 02-Jun-2000 #sequence_revision 02-Jun-2000 #text_change 02-Feb-2001
C:Accession: T49283
R:Rieger, M.; Gabel, C.; Mueller-Auer, S.; Schaefer, M.; Zipp, M.; Mewes, H.W.; Rudd, S.
submitted to the Protein Sequence Database, April 2000
A:Reference number: Z25021
A:Accession: T49283
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-952 <RIB>
A:Cross-references: EMBL:AL132963; GSPDB:GN00061; ATSP:T21J18.140
A:Experimental source: cultivar Columbia; BAC clone T21J18
C:Genetics:
A:Gene: ATSP:T21J18.140
A:Map position: 3
A:Introns: 212/1; 250/3; 353/3; 425/3; 498/2; 523/3; 553/3
C:Superfamily: endopeptidase Clp ATP-binding chain
C:Keywords: ATP; molecular chaperone; nucleotide binding; P-loop
F:323-330/Region: nucleotide-binding motif A (P-loop)
F:390-395/Region: nucleotide-binding motif B
F:666-673/Region: nucleotide-binding motif A (P-loop)
F:734-739/Region: nucleotide-binding motif B
F:329/Binding site: ATP (Lys) #status predicted
F:672/Binding site: ATP (Lys) #status predicted

Query Match 1.5%; Score 8; DB 2; Length 952;
Best Local Similarity 100.0%; Pred. No. 36;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 127 EEIRQSDS 134
| | | | |
Db 381 EEIRQSDS 388

RESULT 25

AE3243
conjugal transfer protein, Dtr system traA [imported] - Agrobacterium tumefaciens (stra
C:Species: Agrobacterium tumefaciens
C>Date: 11-Jan-2002 #sequence_revision 11-Jan-2002 #text_change 18-Nov-2002
C:Accession: AE3243
R:Wood, D.W.; Setubal, J.C.; Kaul, R.; Monks, D.; Chen, L.; Wood, G.E.; Chen, Y.; Woo,
erage, G.; Gillet, W.; Grant, C.; Guenther, D.; Kutyavin, T.; Levy, R.; Li, M.; McClell
; Karp, P.; Romero, P.; Zhang, S.
Science 294, 2317-2323, 2001
A:Authors: Yoo, H.; Tao, Y.; Biddle, P.; Jung, M.; Krespan, W.; Perry, M.; Gordon-Kamm,
star, E.W.
A:Title: The Genome of the Natural Genetic Engineer Agrobacterium tumefaciens C58.
A:Reference number: AB2577; MUID:21608550; PMID:11743193
A:Accession: AE3243
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-1100 <KUR>
A:Cross-references: GB:AE008690; PIDN:AAL46363.1; PID:gl7744153; GSPDB:GN00169
A:Experimental source: strain C58 (Dupont)
C:Genetics:
A:Gene: traA
A:Genome: plasmid

Query Match 1.5%; Score 8; DB 2; Length 1100;
Best Local Similarity 100.0%; Pred. No. 41;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 135 VGDRDHR 142
| | | | |
Db 671 VGDRDHR 678

RESULT 26

T03419
traA protein - Agrobacterium tumefaciens plasmid pTiC58
C:Species: Agrobacterium tumefaciens
C>Date: 24-Mar-1999 #sequence_revision 24-Mar-1999 #text_change 08-Oct-1999
C:Accession: T03419
R:Piper, K.R.; Beck von Bodman, S.; Cook, D.M.; Hwang, I.; Kim, H.; Farrand, S.K.
submitted to the EMBL Data Library, May 1998
A:Reference number: Z14943
A:Accession: T03419
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-1101 <PIP>
A:Cross-references: EMBL:AF010180; NID:g3153171; PIDN:AAC17212.1; PID:g1103910
C:Genetics:
A:Gene: traA
A:Genome: plasmid pTiC58

Query Match 1.5%; Score 8; DB 2; Length 1101;
Best Local Similarity 100.0%; Pred. No. 41;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 135 VGDRDHR 142
| | | | |
Db 672 VGDRDHR 679

RESULT 27

C97702
cell surface antigen [imported] - Rickettsia conorii (strain Malish 7)
C:Species: Rickettsia conorii
C>Date: 30-Sep-2001 #sequence_revision 30-Sep-2001 #text_change 30-Sep-2001
C:Accession: C97702
R:Ogata, H.; Audic, S.; Renesto-Audiffren, P.; Fournier, P.E.; Barbe, V.; Samson, D.; R.
Science 293, 2093-2098, 2001
A:Title: Mechanisms of Evolution in Rickettsia conorii and Rickettsia prowazekii.
A:Reference number: A97700; MUID:2142074; PMID:11557893
A:Accession: C97702
A:Status: preliminary

A:Molecule type: DNA
A:Residues: 1-1902 <KUR>
C:Cross-references: GB:AE006914; PIDN:AAL02557.1; PID:gl51619052; GSPDB:GN00173
C:Genetics:
A:Gene: scal

Query Match
Best Local Similarity 100.0%; Score 8; DB 2; Length 1902;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 405 KIKLEKE 412
|||||
DB 356 KIKLEKE 363

RESULT 28
T18519
myosin X - bovine
C:Species: Bos primigenius taurus (cattle)
C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 02-Feb-2001
C:Accession: T18519
R:Corey, D.P.; Derfler, B.H.; Solc, C.K.; Duyk, G.M.; Cheney, R.E.
submitted to the EMBL Data Library, April 1996
A:Description: Cloning and expression of myosin X, a novel unconventional myosin with pI
A:Reference number: Z18942
A:Accession: T18519
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 1-2052 <COR>
A:Cross-references: EMBL:U55042; NID:gl755048; PID:gl755049; PIDN:AAB39486.1
A:Experimental source: aorta
C:Superfamily: myosin motor domain homology
C:Keywords: nucleotide binding; P-loop
P:66-727/Domain: myosin motor domain homology <MWO>
P:157-164/Region: nucleotide-binding motif A (P-loop)

Query Match
Best Local Similarity 100.0%; Score 8; DB 2; Length 2052;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 248 REEEKRK 255
|||||
DB 824 REEEKRK 831

RESULT 29
S45032
homeoic protein SPOX TAL - Tethya aurantia (fragment)
C:Species: Tethya aurantia
C:Date: 06-Oct-1994 #sequence_revision 26-Jul-1996 #text_change 15-Oct-1999
C:Accession: S45032
R:Degnan, B.M.; Degnan, S.M.; Giusti, A.; Morse, D.E.
submitted to the EMBL Data Library, May 1994
A:Description: A Hox homeobox gene in sponges.
A:Reference number: S45030
A:Accession: S45032
A:Molecule type: DNA
A:Residues: 1-23 <DEG>
A:Cross-references: EMBL:X79265; NID:9488780; PIDN:CAA5852.1; PID:g495148
C:Superfamily: unassigned homeobox proteins; homeobox homology
C:Keywords: DNA binding; homeobox; nucleus; transcription regulation

Query Match
Best Local Similarity 100.0%; Score 7; DB 2; Length 23;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 288 ELAERLK 294
|||||
DB 12 ELAERLK 18

RESULT 30
B91253

hypothetical protein ECs4994 [imported] - Escherichia coli (strain O157:H7, substrain R
C:Species: Escherichia coli
C:Date: 18-Jul-2001 #sequence_revision 18-Jul-2001 #text_change 18-Jul-2001
C:Accession: B91253
R:Hayashi, T.; Makino, M.; Kurokawa, K.; Ishii, K.; Yokoyama, K.; Han, C.G.
sasawara, N.; Yasunaga, T.; Kuhara, S.; Shiba, T.; Hattori, M.; Shinagawa, H.
DNA Res. 8, 11-22, 2001
A:Title: Complete genome sequence of enterohemorrhagic Escherichia coli O157:H7 and gen
A:Reference number: A99629; MUID:21156231; PMID:11258796
A:Accession: B91253
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-60 <HAY>
A:Cross-references: GB:BA000007; PIDN:BA838417.1; PID:gl3364470; GSPDB:GN00154
A:Experimental source: strain O157:H7, substrain RMD 050952
C:Genetics:
A:Gene: ECs4994

Query Match
Best Local Similarity 100.0%; Score 7; DB 2; Length 60;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 494 AQAPSSP 500
|||||
DB 48 AQAPSSP 54

RESULT 31
AG1905
hypothetical protein asl0793 [imported] - Nostoc sp. (strain PCC 7120)
C:Species: Nostoc sp. PCC 7120
A:Note: Nostoc sp. strain PCC 7120 is a synonym of Anabaena sp. strain PCC 7120
C:Date: 14-Dec-2001 #sequence_revision 14-Dec-2001 #text_change 09-Dec-2002
C:Accession: AG1905
R:Kaneko, T.; Makamura, Y.; Wolk, C.P.; Kuritz, T.; Sasamoto, S.; Watanabe, A.; Iriguch
Nakazaki, N.; Shimpo, S.; Sugimoto, M.; Takazawa, M.; Yamada, M.; Tabata, I.
DNA Res. 8, 205-213, 2001
A:Title: Complete Genomic Sequence of the Filamentous Nitrogen-fixing Cyanobacterium An
A:Reference number: AB1807; MUID:21595285; PMID:11759840
A:Accession: AG1905
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-62 <KUR>
A:Cross-references: GB:BA000019; PIDN:BA872750.1; PID:gl7130138; GSPDB:GN00173
A:Experimental source: strain PCC 7120
C:Genetics:
A:Gene: asl0793

Query Match
Best Local Similarity 100.0%; Score 7; DB 2; Length 62;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 471 DLSAGGQ 477
|||||
DB 21 DLSAGGQ 27

RESULT 32
AC1755
hypothetical protein lin2594 [imported] - Listeria innocua (strain Clip11262)
C:Species: Listeria innocua
C:Date: 27-Nov-2001 #sequence_revision 27-Nov-2001 #text_change 27-Nov-2001
C:Accession: AC1755
R:Glaser, P.; Frangeul, L.; Buchrieser, C.; Amend, A.; Baquero, F.; Berche, P.; Bloecke
.; Dominguez-Bernal, G.; Duchaud, E.; Durand, L.; Dussurge, O.; Entian, K.D.; Fsihi, H.
D.; Jones, L.M.; Karst, U.
Science 294, 849-852, 2001
A:Authors: Kreft, J.; Kuhn, M.; Kunst, F.; Kurapkat, G.; Madueno, E.; Maitourram, A.; Ma
ok, C.; Schlueter, T.; Simoes, N.; Tierrez, A.; Vazquez-Boland, J.A.; Voss, H.; Wehland,
A:Title: Comparative genomics of Listeria species.
A:Reference number: AB1077; MUID:21537279; PMID:11679669
A:Accession: AC1755
A:Status: preliminary

A:Molecule type: DNA
A:Residues: 1-75 <GLA>
A:Cross-references: GB:AL592022; PIDN:CAC97811.1; PID:g16415106; GSPDB:GN00178
A:Experimental source: strain Clp11262
C:Genetics:
A:Gene: lin2584

Query Match 1.3%; Score 7; DB 2; Length 75;
Best Local Similarity 100.0%; Pred. No. 34;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 293 LKXLIQ 299
DB 58 LKXLIQ 64
|||||

RESULT 33
D75326
hypothetical protein - Deinococcus radiodurans (strain RL)
C:Species: Deinococcus radiodurans
C:Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 31-Mar-2000
C:Accession: D75326
R:White, O.; Eisen, J.A.; Heidelberg, J.F.; Hickey, E.K.; Peterson, J.D.; Dodson, R.J.;
M.; Shen, M.; Vamathevan, J.J.; Lam, P.; McDonald, L.; Utterback, T.; Zalewski, C.; Ma
S.; Smith, H.O.; Venter, J.C.; Fraser, C.M.
Science 286, 1571-1577, 1999
A:Title: Genome sequence of the radioresistant bacterium Deinococcus radiodurans R1.
A:Reference number: A75250; MUID:20036896; PMID:10567266
A:Accession: D75326
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-79 <WHI>
A:Cross-references: GB:AE002038; GB:AE000513; MID:g6459790; PIDN:AAF11565.1; PID:g645980
A:Experimental source: strain R1
C:Genetics:
A:Gene: DR2007
A:Map position: 1

Query Match 1.3%; Score 7; DB 2; Length 79;
Best Local Similarity 100.0%; Pred. No. 36;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 58 ELRQLE 64
DB 12 ELRQLE 18
|||||

RESULT 34
S78290
ribosomal protein L21, chloroplast - Odontella sinensis chloroplast
C:Species: chloroplast Odontella sinensis
C:Date: 17-Feb-1998 #sequence_revision 26-Feb-1998 #text_change 20-Jun-2000
C:Accession: S78290
R:Kowalik, K.V.; Stoebe, B.; Schaffran, I.; Kroth-Pancic, P.; Freier, U.
Plant Mol. Biol. Rep. 13, 336-342, 1995
A:Title: The Chloroplast Genome of a chlorophyll a+c- containing Alga, Odontella sinensi
A:Reference number: S78238
A:Accession: S78290
A:Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-105 <ROW>
A:Cross-references: EMBL:Z67753; MID:g1185127; PIDN:CAA91663.1; PID:g1185180
A:Note: the nucleotide sequence was submitted to the EMBL Data Library, November 1995
C:Genetics:
A:Gene: rpl21
A:Genome: chloroplast
C:Superfamily: Escherichia coli ribosomal protein L21
C:Keywords: chloroplast; protein biosynthesis; ribosome

Query Match 1.3%; Score 7; DB 2; Length 105;
Best Local Similarity 100.0%; Pred. No. 47;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 152 LKKBTL 158
DB 30 LKKBTL 36
|||||

RESULT 35
S10587
cystatin C - rat
C:Species: Rattus sp. (rat)
C:Date: 21-Nov-1993 #sequence_revision 03-Nov-1995 #text_change 16-Jul-1999
C:Accession: S10587
R:Ennard, F.; Ennard, A.; Faucher, D.; Capony, J.P.; Derancourt, J.; Brillard, M.; Gaut
Biol. Chem. Hoppe-Seyler 371(Suppl.), 161-166, 1990
A:Title: Rat cystatin C: the complete amino acid sequence reveals a site for N-glycosyl
A:Reference number: S10587; MUID:90380276; PMID:2400577
A:Accession: S10587
A:Status: preliminary
A:Molecule type: protein
A:Residues: 1-120 <ESN>
A:Note: 43-Asn was also found
A:Note: the sequence from Fig. 2 is inconsistent with that from Fig. 1 in having 18-Ala
C:Superfamily: cystatin; cystatin homology
P:9-120/Domain: cystatin homology <CYS>

Query Match 1.3%; Score 7; DB 2; Length 120;
Best Local Similarity 100.0%; Pred. No. 53;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 241 REGVQRA 247
DB 20 REGVQRA 26
|||||

RESULT 36
T44473
conserved hypothetical protein tnpG [imported] - Shigella flexneri
C:Species: Shigella flexneri
C:Date: 21-Jan-2000 #sequence_revision 21-Jan-2000 #text_change 21-Jul-2000
C:Accession: T44473
R:Moss, J.B.; Cardozo, T.J.; Zychlinsky, A.; Groisman, E.A.
Mol. Microbiol. 33, 74-83, 1999
A:Title: The selC-associated SHI-2 pathogenicity island of Shigella flexneri.
A:Reference number: Z22779; MUID:99340340; PMID:10411725
A:Accession: T44473
A:Status: preliminary; translated from GB/EMBL/DDBJ
A:Molecule type: DNA
A:Residues: 1-121 <WOS>
A:Cross-references: EMBL:AF141323; MID:g5532445; PIDN:RAD4740.1; PID:g5532456
A:Experimental source: strain MS0T; serotype 5a
C:Genetics:
A:Gene: tnpG
C:Superfamily: Agrobacterium tumefaciens insertion sequence IS1312 hypothetical protein

Query Match 1.3%; Score 7; DB 2; Length 121;
Best Local Similarity 100.0%; Pred. No. 53;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 346 LLKQAVE 352
DB 96 LLKQAVE 102
|||||

RESULT 37
S07085
cystatin C precursor - rat (fragment)
C:Species: Rattus norvegicus (Norway rat)
C:Date: 01-Dec-1993 #sequence_revision 03-Aug-1995 #text_change 16-Jul-1999
C:Accession: S07085; S01337; S21109
R:Cole, T.; Dickson, P.W.; Ennard, F.; Averill, S.; Risbridger, G.P.; Gauthier, F.; Sch
Eur. J. Biochem. 186, 35-42, 1989
A:Title: The cDNA structure and expression analysis of the genes for the cysteine prote
A:Reference number: S07085; MUID:90092122; PMID:2689174
A:Accession: S07085

A>Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-127 <COL>
A:Cross-references: EMBL:X16957; NID:g56041; PIDN:CRAA34831.1; PID:g736290
R:Esnard, A.; Ennard, F.; Paucher, D.; Gauthier, F.
FEBS Lett. 236, 475-478, 1988
A>Title: Two rat homologues of human cystatin C.
A:Reference number: S01337; MUID:88313020; PMID:3044831
A:Accession: S01337
A:Molecule type: protein
A:Residues: 8-49 <PSN>
R:Esnard, A.; Ennard, F.; Guillou, F.; Gauthier, F.
FEBS Lett. 300, 131-135, 1992
A>Title: Production of the cysteine proteinase inhibitor cystatin C by rat Sertoli cells
A:Reference number: S21109; MUID:92225121; PMID:1563513
A:Accession: S21109
A:Molecule type: protein
A:Residues: 8, 'XX', 11-20 <ES2>
C:Superfamily: cystatin; cystatin homology
C:Keywords: cysteine proteinase inhibitor
F:16-127/Domain: cystatin homology <CRS>
F:80-90,104-124/Disulfide bonds: #status predicted

Query Match 1.3%; Score 7; DB 2; Length 127;
Best Local Similarity 100.0%; Pred. No. 56;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 241 EGVQQA 247
|||||
DB 27 EGVQQA 33

RESULT 38
T4497
hypothetical protein [imported] - Thermomonospora fusca
C:Species: Thermomonospora fusca
C>Date: 21-Jan-2000 #sequence_revision 21-Jan-2000 #text_change 21-Jan-2000
C:Accession: T4497
R:Irwin, D.I.; Zhang, S.; Wilson, D.B.
submitted to the EMBL Data Library, April 1999
A:Description: Characterization of a Thermomonospora fusca family 48 exocellulase B6.
A:Reference number: Z22783
A:Accession: T4497
A>Status: preliminary; translated from GB/EMBL/DDB
A:Molecule type: DNA
A:Residues: 1-128 <IRW>
A:Cross-references: EMBL:AF144563; PIDN:AAD39948.1
A:Experimental source: strain YX

Query Match 1.3%; Score 7; DB 2; Length 128;
Best Local Similarity 100.0%; Pred. No. 56;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 SSPGQPE 8
|||||
DB 108 SSPGQPE 114

RESULT 39
S77948
major allergen Par j I (clone Pb) - Parietaria judaica
C:Species: Parietaria judaica
C>Date: 21-Apr-1997 #sequence_revision 18-Jul-1997 #text_change 02-Sep-2000
C:Accession: S77948; S43682; S41944
R:Geraci, D.
submitted to the EMBL Data Library, September 1995
A:Description: cDNA cloning, expression and primary structure of Par j I, a major allergen
A:Reference number: S77948
A:Accession: S77948
A:Molecule type: mRNA
A:Residues: 1-133 <GER>
A:Cross-references: EMBL:X77414; NID:g992611; PIDN:CRAA54587.1; PID:g992612
A:Experimental source: clone Pb

A>Note: this is a revision to the sequence from reference S43682
R:Costa, M.A.; Colombo, P.; Izzo, V.; Kennedy, K.; Venturilla, S.; Cocchiara, R.; Mistr
FEBS Lett. 341, 182-186, 1994
A>Title: cDNA cloning, expression and primary structure of Par j I, a major allergen of
A:Reference number: S43682; MUID:94185791; PMID:8137937
A:Accession: S43682
A:Molecule type: mRNA
A:Residues: 1-110, 'RSRPPTKHWDRP', 124, 'LSFRP', 130, 'HRKKPNDAFSTLG' <COS>
A:Cross-references: EMBL:X77414
A>Note: the authors translated the codon GTT for residue 88 as Leu
A>Note: this sequence has been revised in reference S77948
C:Genetics:
A:Gene: Pmai
C:Superfamily: phospholipid transfer protein

Query Match 1.3%; Score 7; DB 2; Length 133;
Best Local Similarity 100.0%; Pred. No. 58;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 116 EKEPSKG 122
|||||
DB 16 EKEPSKG 22

RESULT 40
B65240
hypothetical protein, 13.4K - Escherichia coli (strain K-12) insertion sequence IS2
C:Species: Escherichia coli
C>Date: 12-Sep-1997 #sequence_revision 23-Sep-1997 #text_change 01-Mar-2002
C:Accession: B65240; F64891; H64763; B65092; S56497; JQ0039
R:Blattner, F.R.; Plunkett III, G.; Bloch, C.A.; Perna, N.T.; Burland, V.; Riley, M.; C
A.; Rose, D.J.; Mau, B.; Shao, Y.
Science 277, 1453-1462, 1997
A>Title: The complete genome sequence of Escherichia coli K-12.
A:Reference number: A64720; MUID:97426617; PMID:9278503
A:Accession: B65240
A>Status: nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-136 <BLAT>
A:Cross-references: GB:AB000498; GB:U00096; NID:g2367368; PIDN:AAC77228.1; PID:g1790723
A:Experimental source: strain K-12, substrain MG1655
A:Genetics: Y21
A:Accession: F64891
A>Status: nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-136 <BLA2>
A:Cross-references: GB:AE000237; GB:U00096; NID:g1787665; PIDN:AAC74485.1; PID:g1787670
A:Experimental source: strain K-12, substrain MG1655
A:Genetics: Y22
A:Accession: H64763
A>Status: nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-136 <BLA3>
A:Cross-references: GB:AE000143; GB:U00096; NID:g1786554; PIDN:AAC73463.1; PID:g1786557
A:Experimental source: strain K-12, substrain MG1655
A:Genetics: Y11
A:Accession: B65092
A>Status: nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 16-136 <BLA4>
A:Cross-references: GB:AB000386; GB:U00096; NID:g2367187; PIDN:AAC76080.1; PID:g1789423
A:Experimental source: strain K-12, substrain MG1655
A:Genetics: Y25
A:Accession: F65069
A>Status: nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 16-136 <BLA5>
A:Cross-references: GB:AB000369; GB:U00096; NID:g2367168; PIDN:AAC75900.1; PID:g1789225,
A:Experimental source: strain K-12, substrain MG1655
A:Genetics: Y24
R:Burland, V.; Plunkett III, G.; Sofia, H.J.; Daniels, D.L.; Blattner, F.R.
Nucleic Acids Res. 23, 2105-2119, 1995
A>Title: Analysis of the Escherichia coli genome VI: DNA sequence of the region from 92

A:Reference number: S56314; MUID:95334362; PMID:7610040
A:Accession: S56497
A>Status: nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-136 <BUR>
A:Cross-references: EMBL:U14003; NID:g1263172; PIDN:AAA97168.1; PID:g537113
A:Genetics: Y26
A>Note: the nucleotide sequence was submitted to the EMBL Data Library, August 1994
R:Ronecker, H.J.; Rak, B.
Gene 59, 291-296, 1987
A:Title: Genetic organization of insertion element IS2 based on a revised nucleotide sequence
A:Reference number: A91582; MUID:98137965; PMID:2630172
A:Accession: JQ0039
A>Status: translation not shown
A:Molecule type: DNA
A:Residues: 16-136 <RON>
A:Cross-references: GB:V00279; GB:J01732; GB:M18426; NID:g41525; PIDN:CAA23542.1; PID:g41525
A:Experimental source: K12, strain HB101
A:Genetics: HB1
C:Genetics: <Y21>
A:Gene: Y121
A:Mobile element: insertion sequence IS2
C:Genetics: <Y22>
A:Gene: Y121_2
A:Mobile element: insertion sequence IS2
C:Genetics: <Y11>
A:Gene: Y121_1
A:Mobile element: insertion sequence IS2
C:Genetics: <Y25>
A:Gene: Y121_5
A:Mobile element: insertion sequence IS2
C:Genetics: <Y24>
A:Gene: Y121_4
A:Mobile element: insertion sequence IS2
C:Genetics: <Y26>
A:Mobile element: insertion sequence IS2
C:Genetics: <HB1>
A:Mobile element: insertion sequence IS2
A:Start codon: GAG
C:Superfamily: Agrobacterium tumefaciens insertion sequence IS1312 hypothetical protein

Query Match 1.3%; Score 7; DB 2; Length 136;
Best Local Similarity 100.0%; Pred. No. 59;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 346 LLKEAVE 352
DB 111 LLKEAVE 117
|||||

RESULT 41
D64964
hypothetical protein - Escherichia coli (strain K-12) insertion sequence IS2
C:Species: Escherichia coli
C:Date: 12-Sep-1997 #sequence_revision 17-Sep-1997 #text_change 01-Mar-2002
C:Accession: D64964
R:Blattner, F.R.; Plunkett III, G.; Bloch, C.A.; Perna, N.T.; Burland, V.; Riley, M.; Cohen, D.; Rose, D.J.; Mau, B.; Shao, Y.
Science 277, 1453-1462, 1997
A:Title: The complete genome sequence of Escherichia coli K-12.
A:Reference number: A64720; MUID:97426617; PMID:9278503
A:Accession: D64964
A>Status: nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-136 <BLAT>
A:Cross-references: GB:AE000291; GB:U00096; NID:g1788298; PIDN:AACT5058.1; PID:g1788306;
A:Experimental source: strain K-12, substrain MG1655
C:Genetics:
A:Gene: Y121_3
A:Mobile element: insertion sequence IS2
C:Superfamily: Agrobacterium tumefaciens insertion sequence IS1312 hypothetical protein

Query Match 1.3%; Score 7; DB 2; Length 136;

Best Local Similarity 100.0%; Pred. No. 59;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 346 LLKEAVE 352
DB 111 LLKEAVE 117
|||||

RESULT 42
S52933
major allergen Par j (clone PA2) - Parietaria judaica
C:Species: Parietaria judaica
C:Date: 19-Mar-1997 #sequence_revision 19-Mar-1997 #text_change 02-Sep-2000
C:Accession: S52933
R:Geraci, D.
submitted to the EMBL Data Library, February 1995
A:Reference number: S52933
A:Accession: S52933
A:Molecule type: mRNA
A:Residues: 1-138 <GER>
A:Cross-references: EMBL:X85012; NID:g706810; PIDN:CAA59370.1; PID:g706811
A:Experimental source: clone PA2
C:Superfamily: phospholipid transfer protein

Query Match 1.3%; Score 7; DB 2; Length 138;
Best Local Similarity 100.0%; Pred. No. 60;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 116 EKEPSKG 122
DB 58 EKEPSKG 64
|||||

RESULT 43
A90258
conserved hypothetical protein [imported] - Sulfolobus solfataricus
C:Species: Sulfolobus solfataricus
C:Date: 24-May-2001 #sequence_revision 24-May-2001 #text_change 24-May-2001
C:Accession: A90258
R:She, Q.; Singh, R.K.; Confalonieri, F.; Zivanovic, Y.; Allard, G.; Awayez, M.J.; Chan
Jong, I.; Jeffries, A.C.; Kozera, C.J.; Medina, N.; Peng, X.; Thi-Ngoc, H.P.; Redder,
arrett, R.A.; Ragan, M.A.; Sensen, C.W.; Van der Oost, J.
submitted to GenBank, April 2001
A:Description: Sulfolobus solfataricus complete genome.
A:Reference number: A99139
A:Accession: A90258
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-138 <KUR>
A:Cross-references: GB:AE006641; NID:g13814244; PIDN:AAK41320.1; GSPDB:GN00155
C:Genetics:
A:Gene: SSO1056

Query Match 1.3%; Score 7; DB 2; Length 138;
Best Local Similarity 100.0%; Pred. No. 60;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 332 LKAEER 338
DB 111 LKAEER 117
|||||

RESULT 44
G98112
hypothetical protein spr1930 [imported] - Streptococcus pneumoniae (strain R6)
C:Species: Streptococcus pneumoniae
C:Date: 22-Oct-2001 #sequence_revision 22-Oct-2001 #text_change 22-Oct-2001
C:Accession: G98112
R:Hoskins, J.A.; Alborn Jr., W.; Arnold, J.; Blaszcak, L.; Burgett, S.; DeHoff, B.S.;
e, R.; LeBlanc, D.J.; Lee, L.N.; Lefkowitz, E.J.; Lu, J.; Matsushima, P.; McAnen, S.;
J. P.; Sun, P.M.; Winkler, M.E.
Y. Bacteriol. 183, 5709-5717, 2001
A:Authors: Yang, Y.; Young-Bellido, M.; Zhao, G.; Zook, C.; Baltz, R.H.; Jaskunas, S.R.

A:Title: Genome of the Bacterium Streptococcus pneumoniae Strain R6.
A:Reference number: A97872; MUID:21429245; PMID:11544234
A:Accession: G98112
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-154 <KUR>
A:Cross-references: GB:AE007317; PIDN:AAL00732.1; PID:G15459627; GSPDB:GN00174
C:Genetics:
A:Gene: spr1930

Query Match 1.3%; Score 7; DB 2; Length 154;
Best Local Similarity 100.0%; Pred. No. 67;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 404 KKIKKLE 410
DB 69 KKIKKLE 75

RESULT 45
B95248
Hypothetical protein SP2120 [imported] - Streptococcus pneumoniae (strain TIGR4)
C:Species: Streptococcus pneumoniae
C:Date: 03-Aug-2001 #sequence_revision 03-Aug-2001 #text_change 03-Aug-2001
C:Accession: B95248
R:Tetreltin, H.; Nelson, K.E.; Paulsen, I.T.; Eisen, J.A.; Read, T.D.; Peterson, S.; Heid
on, J.D.; Unayam, L.A.; White, O.; Salzberg, S.L.; Lewis, M.R.; Radune, D.; Holtzapple,
Nelson, I.; Hackey, E.K.; Holt, I.E.
A:Authors: Loftus, B.J.; Yang, F.; Smith, H.O.; Venter, J.C.; Dougherty, B.A.; Morrison,
A:Title: Complete Genome Sequence of a virulent isolate of Streptococcus pneumoniae.
A:Reference number: A95000; MUID:21357209; PMID:11463916
A:Accession: B95248
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-155 <KUR>
A:Cross-references: GB:AE005672; PIDN:AAK76179.1; PID:G14973632; GSPDB:GN00164; TIGR:SP4
A:Experimental source: strain TIGR4
C:Genetics:
A:Gene: SP2120

Query Match 1.3%; Score 7; DB 2; Length 155;
Best Local Similarity 100.0%; Pred. No. 67;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 404 KKIKKLE 410
DB 70 KKIKKLE 76

RESULT 46
T49567
related to attachment protein [imported] - Neurospora crassa
N:Alternate names: protein B208.150
C:Species: Neurospora crassa
C:Date: 02-Jun-2000 #sequence_revision 02-Jun-2000 #text_change 02-Jun-2000
C:Accession: T49567
R:Schulte, U.; Aign, V.; Hoheisel, J.; Brandt, P.; Fartmann, B.; Holland, R.; Nyakatura,
submitted to the Protein Sequence Database, May 2000
A:Reference number: Z25022
A:Accession: T49567
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-158 <SCH>
A:Cross-references: EMBL:AL355930; GSPDB:GN00116; NCSP:B208.150
A:Experimental source: BAC clone B208; strain OR74A
C:Genetics:
A:Gene: NCSP:B208.150
A:Map position: 6

Query Match 1.3%; Score 7; DB 2; Length 158;
Best Local Similarity 100.0%; Pred. No. 68;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 240 KEEGVQR 246
DB 62 KEEGVQR 68

RESULT 47
T32043
Hypothetical protein K07E8.6 - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C:Date: 29-Oct-1999 #sequence_revision 29-Oct-1999 #text_change 29-Oct-1999
C:Accession: T32043
R:Jones, K.; Kramer, J.
A:Submitted to the EMBL Data Library, July 1997
A:Description: The sequence of C. elegans cosmid K07E8.
A:Reference number: Z21116
A:Accession: T32043
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-159 <JON>
A:Cross-references: EMBL:AF016678; PIDN:AAB66151.1; GSPDB:GN00020; CESP:K07E8.6
A:Experimental source: strain Bristol N2; clone K07E8
C:Genetics:
A:Gene: CESP:K07E8.6
A:Map position: 2

Query Match 1.3%; Score 7; DB 2; Length 159;
Best Local Similarity 100.0%; Pred. No. 69;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 170 EEKLAAL 176
DB 54 EEKLAAL 60

RESULT 48
E71183
VPS29-like phosphoesterase-related protein PH1746 [similarity] - Pyrococcus horikoshii
C:Species: Pyrococcus horikoshii
C:Date: 14-Aug-1998 #sequence_revision 14-Aug-1998 #text_change 24-Aug-2001
C:Accession: E71183
R:Kawarabayashi, Y.; Sawada, M.; Horikawa, H.; Haikawa, Y.; Hino, Y.; Yamamoto, S.; Seki,
M.; Ohfuku, Y.; Funahashi, T.; Tanaka, T.; Kudoh, Y.; Yamazaki, J.; Kushida, N.; Oguch,
DNA Res. 5: 55-76, 1998
A:Title: Complete sequence and gene organization of the genome of a hyper-thermophilic
A:Reference number: A71000; MUID:98344137; PMID:9679194
A:Accession: E71183
A:Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-163 <KAW>
A:Cross-references: GB:AP000007; MUID:G3236134; PIDN:BAA30860.1; PID:G3258177
A:Experimental source: strain OT3
A:Note: this accession replaces an interim accession for a sequence replaced by GenBank
C:Genetics:
A:Gene: PH1746
C:Superfamily: human vacuolar protein-sorting protein VPS29 homolog; phosphoesterase co

Query Match 1.3%; Score 7; DB 2; Length 163;
Best Local Similarity 100.0%; Pred. No. 70;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 153 GKEITLL 159
DB 123 GKEITLL 129

RESULT 49
A69454
Hypothetical protein AF1634 - Archaeoglobus fulgidus
C:Species: Archaeoglobus fulgidus
C:Date: 05-Dec-1997 #sequence_revision 05-Dec-1997 #text_change 22-Oct-1999
C:Accession: A69454
R:Klenk, H.P.; Clayton, R.A.; Tomb, J.F.; White, O.; Nelson, K.E.; Ketchum, K.A.; Dodson

```

.; Fleischmann, R.D.; Quackenbush, J.; Lee, N.H.; Sutton, G.G.; Gill, S.; Kirkness, E.F.;
Glodek, A.; Zhou, L.; Overbeek, R.; Gocayne, J.D.; Weidman, J.F.; McDonald, L.
Nature 390, 364-370, 1997
A:Authors: Uterback, T.; Cotton, M.D.; Spriggs, T.; Artiach, P.; Kaine, B.P.; Sykes, S.
Smith, H.O.; Woese, C.R.; Venter, J.C.
A:Title: The complete genome sequence of the hyperthermophilic, sulfate-reducing archae
A:Reference number: A69250; MUID:98049343; PMID:9389475
A:Accession: A69454
A:Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-172 <KLE>
A:Cross-references: GB:AE000989; GB:AE000782; NID:G2689312; PIDN:AA89617.1; PID:G264892

Query Match      1.3%; Score 7; DB 2; Length 172;
Best Local Similarity 100.0%; Pred. No. 74;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 153 GKXITLL 159
DB 78 GKXITLL 84
|||||

RESULT 50
AE2913
Conserved hypothetical protein Atu2742 [imported] - Agrobacterium tumefaciens (strain C5
A:Species: Agrobacterium tumefaciens
C:Date: 11-Jan-2002 #sequence_revision 11-Jan-2002 #text_change 18-Nov-2002
C:Accession: AE2913
R:Wood, D.W.; Setubal, J.C.; Kaul, R.; Monks, D.; Chen, L.; Wood, G.E.; Chen, Y.; Woo, I
erage, G.; Gillet, W.; Grant, C.; Guenther, D.; Kutyavin, T.; Levy, R.; Li, M.; McClell
; Karp, P.; Romero, P.; Zhang, S.
Science 294, 2317-2323, 2001
A:Authors: Yoo, H.; Tao, Y.; Biddle, P.; Jung, M.; Krespan, W.; Perry, M.; Gordon-Kamm,
ster, B.W.
A:Title: The Genome of the Natural Genetic Engineer Agrobacterium tumefaciens C58.
A:Reference number: AB2577; MUID:21608550; PMID:11743193
A:Accession: AE2913
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-176 <KUR>
A:Cross-references: GB:AE008698; PIDN:AAL43723.1; PID:gl7741254; GSPDB:GN00186
A:Experimental source: strain C58 (Dupont)
A:Genetics:
A:Gene: Atu2742
A:Map position: circular chromosome

Query Match      1.3%; Score 7; DB 2; Length 176;
Best Local Similarity 100.0%; Pred. No. 75;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 55 VSEELSR 61
DB 43 VSEELSR 49
|||||

RESULT 51
T37212
hypothetical protein Y57G7A.2 - Caenorhabditis elegans
A:Species: Caenorhabditis elegans
C:Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 18-Aug-2000
C:Accession: T37212
R:Ozersky, P.
submitted to the EMBL Data Library, March 1999
A:Description: The sequence of C. elegans cosmid Y57G7A.
A:Reference number: Z21634
A:Accession: T37212
A:Status: preliminary; translated from GB/EMBL/DDBJ
A:Molecule type: DNA
A:Residues: 1-177 <OZE>
A:Cross-references: EMBL:AF077542; PIDN:AAC26297.1
A:Genetics:
A:Map position: 2
A:introns: 105/1; 133/1

```

```

A:Note: Y57G7A.2
C:Superfamily: Caenorhabditis elegans hypothetical protein Y57G7A.2

Query Match      1.3%; Score 7; DB 2; Length 177;
Best Local Similarity 100.0%; Pred. No. 76;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 424 NKALLEM 430
DB 109 NKALLEM 115
|||||

RESULT 52
C85165
hypothetical protein dl3556w [imported] - Arabidopsis thaliana
A:Species: Arabidopsis thaliana (mouse-ear cress)
C:Date: 16-Feb-2001 #sequence_revision 16-Feb-2001 #text_change 16-Feb-2001
C:Accession: C85165
R:Anonymous, The European Union Arabidopsis Genome Sequencing Consortium, The Cold Spri
Nature 402, 769-777, 1999
A:Title: Sequence and analysis of chromosome 4 of the plant Arabidopsis thaliana.
A:Reference number: AB5001; MUID:20083488; PMID:10617198
A:Accession: C85165
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-179 <STO>
A:Cross-references: GB:NC_001268; NID:G5302779; PIDN:CAB46057.1; GSPDB:GN00140
C:Genetics:
A:Gene: dl3556w
A:Map position: 4

Query Match      1.3%; Score 7; DB 2; Length 179;
Best Local Similarity 100.0%; Pred. No. 77;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 250 BEEKKEK 256
DB 108 BEEKKEK 114
|||||

RESULT 53
AB3445
H+-transporting two-sector ATPase (EC 3.6.3.14) [imported] - Brucella melitensis (strai
A:Species: Brucella melitensis
C:Date: 01-Feb-2002 #sequence_revision 01-Feb-2002 #text_change 03-Jun-2002
C:Accession: AB3445
R:DeIvecchio, V.G.; Kaputal, V.; Redkar, R.J.; Patra, G.; Mujer, C.; Los, T.; Ivanova,
.; Mazur, M.; Goltzman, E.; Selkov, E.; Elzer, P.H.; Hagius, S.; O'Callaghan, D.; Letes
Proc. Natl. Acad. Sci. U.S.A. 99, 443-448, 2002
A:Title: The genome sequence of the facultative intracellular pathogen Brucella meliten
A:Reference number: AD3252; PMID:11756688
A:Accession: AB3445
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-180 <KUR>
A:Cross-references: GB:AB008917; PIDN:AAL52725.1; PID:gl7983555; GSPDB:GN00190
A:Experimental source: strain 16M
A:Genetics:
A:Gene: EMBL1544
A:Map position: 1
C:Keywords: hydrolase

Query Match      1.3%; Score 7; DB 2; Length 180;
Best Local Similarity 100.0%; Pred. No. 77;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 332 LKAEER 338
DB 125 LKAEER 131
|||||

RESULT 54
E70209

```

C;Accession: T45013
R;Metcalfe, W.W.; Zhang, J.K.; Apolinario, E.; Sowers, K.R.; Wolfe, R.S.
Proc. Natl. Acad. Sci. U.S.A. 94, 2626-2631, 1997
A;Title: A genetic system for Archaea of the Genus Methanosarcina: liposome-mediated tr
A;Reference number: Z22897; MUID:97226004; PMID:9122246
A;Accession: T45013
A;Status: preliminary; translated from GB/EMBL/DDDBJ
A;Molecule type: DNA
A;Residues: 1-190 <MET>
A;Cross-references: EMBL:U78295; NID:gl763609; PIDN:AA839747.1; PID:gl763613
A;Experimental source: strain C2A
C;Genetics:
A;Genome: plasmid pc2A

Query Match 1.3%; Score 7; DB 2; Length 190;
Best Local Similarity 100.0%; Pred. No. 81;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 SSPGQPS 8
|||
DB 133 SSPGQPS 139
|||

RESULT 57
A81149
outer membrane lipoprotein LolB, probable NMB0873 [imported] - Neisseria meningitidis
C;Species: Neisseria meningitidis
C;Date: 31-Mar-2000 #sequence_revision 31-Mar-2000 #text_change 19-Jan-2001
C;Accession: A81149
R;Tettelin, H.; Saunders, N.J.; Heidelberg, J.; Jeffries, A.C.; Nelson, K.E.; Eisen, J.A.;
Hickey, E.K.; Hatt, D.H.; Salzberg, S.L.; White, C.; Fleischmann, R.D.; Dougherty, B.A.;
ri, H.; Qin, H.; Vamathevan, J.; Gill, J.; Scarlato, V.; Masiagnani, V.; Pizza, M.
Science 287, 1809-1815, 2000
A;Authors: Grandi, G.; Sun, L.; Smith, H.O.; Fraser, C.M.; Moxon, E.R.; Rappuoli, R.; V
A;Title: Complete genome sequence of Neisseria meningitidis serogroup B strain MC58.
A;Reference number: A81000; MUID:2015755; PMID:10710307
A;Accession: A81149
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-193 <P>
A;Cross-references: GB:AE002439; GB:AE002098; NID:g7226100; PIDN:AA41284.1; PID:g72261
A;Experimental source: serogroup B, strain MC58
C;Genetics:
A;Gene: NMB0873

Query Match 1.3%; Score 7; DB 2; Length 193;
Best Local Similarity 100.0%; Pred. No. 82;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 57 EELSROL 63
|||
DB 102 EELSROL 108
|||

RESULT 58
H81874
probable outer membrane lipoprotein NMA1091 [imported] - Neisseria meningitidis (strain
C;Species: Neisseria meningitidis
C;Date: 05-May-2000 #sequence_revision 05-May-2000 #text_change 02-Feb-2001
C;Accession: H81874
R;Parkhill, J.; Achtman, M.; James, K.D.; Bentley, S.D.; Churcher, C.; Klee, S.R.; More
; Holroyd, S.; Jagsels, K.; Leather, S.; Moule, S.; Mungall, K.; Quail, M.A.; Rajandream
Nature 404, 502-506, 2000
A;Title: Complete DNA sequence of a serogroup A strain of Neisseria meningitidis Z2491.
A;Reference number: A81775; MUID:20222556; PMID:10761919
A;Accession: H81874
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-193 <P>
A;Cross-references: GB:AL162755; GB:AL157959; NID:g7379742; PIDN:CA884354.1; PID:g73797
A;Experimental source: serogroup A, strain Z2491
C;Genetics:
A;Gene: NMA1091

```
Query Match      1.3%; Score 7; DB 2; Length 193;
Best Local Similarity 100.0%; Pred. No. 83;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 57 EELSRQL 63
DB 102 EELSRQL 108
      |||||
RESULT 59
SAVLH1
delta large antigen - hepatitis delta virus
N/Alternate names: HDag
C/Species: hepatitis delta virus
A/Note: host Homo sapiens (man)
C/Date: 31-Dec-1992 #sequence_revision 31-Dec-1992 #text_change 16-Jun-2000
C/Accession: A35219
R/Saldanha, J.A.; Thomas, H.C.; Monjardino, J.P.
J. Gen. Virol. 71, 1603-1606, 1990
A/Title: Cloning and sequencing of RNA of hepatitis delta virus isolated from human serum
A/Reference number: A35219; MUID:90324949; PMID:2374010
A/Accession: A35219
A/Molecule type: genomic RNA
A/Residues: 1-195 <SAL>
A/Cross-references: GB:D01075; NID:G221689; PIDN:BA00874.1; PID:G221690
C/Comment: This virus is a replication-defective hepatitis B virus.
C/Superfamily: hepatitis delta virus large antigen
C/Keywords: core protein
F/2-195/Product: delta large antigen #status predicted <MAT>

Query Match      1.3%; Score 7; DB 1; Length 195;
Best Local Similarity 100.0%; Pred. No. 83;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 404 KKIKKLE 410
DB 39 KKIKKLE 45
      |||||
RESULT 60
SAVLH1
delta large antigen - hepatitis delta virus (strain Lebanon-1)
N/Alternate names: HDag
C/Species: hepatitis delta virus
A/Note: host Homo sapiens (man)
C/Date: 31-Dec-1992 #sequence_revision 31-Dec-1992 #text_change 01-Aug-1997
C/Accession: A40247
R/Lee, C.W.; Bih, F.Y.; Chao, Y.C.; Govindarajan, S.; Lai, M.M.C.
Virology 188, 265-273, 1992
A/Title: Evolution of hepatitis delta virus RNA during chronic infection.
A/Reference number: A40247; MUID:92230225; PMID:1566577
A/Accession: A40247
A/Molecule type: genomic RNA
A/Residues: 1-195 <LEE>
A/Cross-references: GB:M84917
C/Comment: This virus is a replication-defective hepatitis B virus.
C/Superfamily: hepatitis delta virus large antigen
C/Keywords: core protein
F/2-195/Product: delta large antigen #status predicted <MAT>

Query Match      1.3%; Score 7; DB 1; Length 195;
Best Local Similarity 100.0%; Pred. No. 83;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 404 KKIKKLE 410
DB 39 KKIKKLE 45
      |||||
RESULT 61
SAVLDM
delta large antigen - hepatitis delta virus (strain Japanese M-1)
```

```
N/Alternate names: HDag
C/Species: hepatitis delta virus
C/Date: 31-Dec-1991 #sequence_revision 31-Dec-1991 #text_change 16-Jun-2000
C/Accession: A36409
R/Imazeki, F.; Omata, M.; Ohto, M.
J. Virol. 64, 5594-5599, 1990
A/Title: Heterogeneity and evolution rates of delta virus RNA sequences.
A/Reference number: A36409; MUID:91012805; PMID:2214027
A/Accession: A36409
A/Molecule type: genomic RNA
A/Residues: 1-195 <IMA>
A/Cross-references: GB:D90190; GB:M58299; NID:G221691; PIDN:BA014214.1; PID:G221692
C/Superfamily: hepatitis delta virus large antigen
F/2-195/Product: delta large antigen #status predicted <MAT>

Query Match      1.3%; Score 7; DB 1; Length 195;
Best Local Similarity 100.0%; Pred. No. 83;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 404 KKIKKLE 410
DB 39 KKIKKLE 45
      |||||
RESULT 62
SS3112
delta antigen - hepatitis delta virus (patient A)
C/Species: hepatitis delta virus
A/Variety: patient A
C/Date: 08-Jul-1995 #sequence_revision 01-Nov-1996 #text_change 26-Aug-1999
R/Lai, M.E.; Mazzoleni, A.P.; Porru, A.; Balestrieri, A.
submitted to the EMBL Data Library, March 1995
A/Reference number: SS3112
A/Accession: SS3112
A/Molecule type: genomic RNA
A/Residues: 1-195 <LAI>
A/Cross-references: EMBL:X85253; NID:G732708; PIDN:CA059509.1; PID:G732709
A/Experimental source: patient A
A/Note: the source is designated as hepatitis D virus
C/Superfamily: hepatitis delta virus large antigen
C/Keywords: core protein
F/2-195/Product: delta large antigen #status predicted <MAT>

Query Match      1.3%; Score 7; DB 2; Length 195;
Best Local Similarity 100.0%; Pred. No. 83;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 404 KKIKKLE 410
DB 39 KKIKKLE 45
      |||||
RESULT 63
G75028
h-transferring ATP synthase, chain E (atpe) PAB1182 - Pyrococcus abyssi (strain Orsay)
C/Species: Pyrococcus abyssi
C/Date: 20-Aug-1999 #sequence_revision 20-Aug-1999 #text_change 15-Sep-2003
C/Accession: G75028
R/anonymous, Genoscope
submitted to the EMBL Data Library, July 1999
A/Description: Pyrococcus abyssi genome sequence: insights into archaeal chromosome str
A/Reference number: A75001
A/Accession: G75028
A/Status: preliminary
A/Molecule type: DNA
A/Residues: 1-199 <KAW>
A/Cross-references: GB:AJ248288; GB:AL096836; NID:G5458960; PIDN:CA050669.1; PID:G545911
A/Experimental source: strain Orsay
C/Genetics:
A/Gene: PAB1182
C/Superfamily: H+-transporting two-sector ATPase
```

```
Query Match      1.3%; Score 7; DB 2; Length 199;
Best Local Similarity 100.0%; Pred. No. 85;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 346 LLKEAVE 352
    |||||
Db 109 LLKEAVE 115

RESULT 64
G95852
conserved hypothetical protein [imported] - Sinorhizobium meliloti (strain 1021) magapla
C:Species: Sinorhizobium meliloti
C:Date: 24-Aug-2001 #sequence_revision 24-Aug-2001 #text_change 30-Sep-2001
C:Accession: G95852
R:Finan, T.M.; Weidner, S.; Wong, K.; Buhrmester, J.; Chain, P.; Vorholter, F.J.; Hernan
Proc. Natl. Acad. Sci. U.S.A. 98, 9889-9894, 2001
A:Title: The complete sequence of the 1,683-kb pSymB megaplasmid from the N2-fixing endo
A:Reference number: A95842; MUID:21396508; PMID:11481431
A:Accession: G95852
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-201 <XUR>
A:CROSS-references: GB:AL5913985; PIDN:CAC48487.1; PID:G15139959; GSPDB:GN00167
A:Experimental source: strain 1021, megaplasmid pSymB
R:Galibert, F.; Finan, T.M.; Long, S.R.; Puhler, A.; Abola, P.; Ampe, F.; Barloy-Hubler,
Pela, D.; Chain, P.; Cowie, A.; Davis, R.W.; Dreano, S.; Federspiel, N.A.; Fisher, R.P.;
L.; Hyman, R.W.; Jones, T.
Science 293, 668-672, 2001
A:Authors: Kahn, D.; Kahn, M.L.; Kalman, S.; Keating, D.H.; Kiss, E.; Komp, C.; Lelaure,
hebaull, P.; Vandenbol, M.; Vorholter, F.J.; Weidner, S.; Wells, D.H.; Wong, K.; Yeh, K.
A:Title: The composite genome of the legume symbiont Sinorhizobium meliloti.
A:Reference number: A96039; MUID:21368234; PMID:11474104
A:Contents: annotation
C:Genetics:
A:Gene: SMD20087
A:Genome: plasmid

Query Match      1.3%; Score 7; DB 2; Length 201;
Best Local Similarity 100.0%; Pred. No. 85;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 478 GSUTDSG 484
    |||||
Db 112 GSUTDSG 118

RESULT 65
C69375
hypothetical protein AF1003 - Archaeoglobus fulgidus
C:Species: Archaeoglobus fulgidus
C:Date: 05-Dec-1997 #sequence_revision 05-Dec-1997 #text_change 22-Oct-1999
C:Accession: C69375
R:Klenk, H.P.; Clayton, R.A.; Tomb, J.F.; White, O.; Nelson, K.E.; Ketchum, K.A.; Dodson
F.; Fleischmann, R.D.; Quackenbush, J.; Lee, N.H.; Sutton, G.G.; Gill, S.; Kirkness, E.F.;
Glodek, A.; Zhou, L.; Overbeek, R.; Gocayne, J.D.; Weidman, J.F.; McDonald, L.
Nature 390, 364-370, 1997
A:Authors: Uitterback, T.; Cotton, M.D.; Spriggs, T.; Artach, P.; Kaine, B.P.; Sykes, S.
Smith, H.O.; Woese, C.R.; Venter, J.C.
A:Title: The complete genome sequence of the hyperthermophilic, sulfate-reducing archaeo
A:Reference number: A69250; MUID:98049343; PMID:9389475
A:Accession: C69375
A>Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-211 <MLE>
A:CROSS-references: GB:AE001034; GB:AE000782; NID:G2689357; PIDN:AAB90239.1; PID:G264959

Query Match      1.3%; Score 7; DB 2; Length 211;
Best Local Similarity 100.0%; Pred. No. 89;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 165 TLSTPEE 171
```

```
Db 33 TLSTPEE 39
    |||||
RESULT 66
SAVLDN
delta large antigen - hepatitis delta virus (strain Nauru)
N:Alternate names: HDAG
C:Species: hepatitis delta virus
C:Date: 31-Dec-1991 #sequence_revision 31-Dec-1991 #text_change 16-Jul-1999
C:Accession: A36212
R:Chao, Y.C.; Chang, M.F.; Gust, I.; Lai, M.M.C.
Virolgy 178, 384-392, 1990
A:Title: Sequence conservation and divergence of hepatitis delta virus RNA.
A:Reference number: A36212; MUID:91020976; PMID:2219700
A:Accession: A36212
A:Molecule type: genomic RNA
A:Residues: 1-214 <CHA>
A:CROSS-references: EMBL:IM34325; EMBL:IM58629; NID:G329994; PIDN:AAB59753.1; PID:G329995
R:Otto, J.C.; Casey, P.J.
J. Biol. Chem. 271, 4569-4572, 1996
A:Title: The hepatitis delta virus large antigen is farnesylated both in vitro and in a
A:Reference number: A58234; MUID:96214864; PMID:8617711
A:Contents: annotation; confirmation of farnesylation rather than geranyl-geranylation
C:Superfamily: hepatitis delta virus large antigen
C:Keywords: core protein; lipoprotein; methylated carboxyl end; prenylated cysteine
P:2-211/Product: delta large antigen #status predicted <MAR>
P:211/Binding site: farnesyl (Cys) (covalent) #status experimental
P:211/Modified site: methyl ester carboxyl end (Cys) (in mature form) #status predicted

Query Match      1.3%; Score 7; DB 1; Length 214;
Best Local Similarity 100.0%; Pred. No. 91;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 404 KKIKKLE 410
    |||||
Db 39 KKIKKLE 45

RESULT 67
A53175
delta large antigen - hepatitis delta virus (strain Taiwan)
N:Alternate names: HDAG
C:Species: hepatitis delta virus
C:Date: 12-May-1994 #sequence_revision 12-May-1994 #text_change 09-Jun-2000
C:Accession: A53175; PD0009
R:Chao, Y.C.; Lee, C.M.; Tang, H.S.; Govindarajan, S.; Lai, M.M.C.
Hepatology 13, 345-352, 1991
A:Title: Molecular cloning and characterization of an isolate of hepatitis delta virus :
A:Reference number: A53175; MUID:91139069; PMID:1995441
A:Accession: A53175
A>Status: preliminary
A:Molecule type: genomic RNA
A:Residues: 1-214 <CHA>
A:CROSS-references: GB:M36590; NID:G329991
R:Tan, W.J.; Zhan, M.Y.; Miao, J.; Cong, X.
Chinese J. Virol. 12, 307-316, 1996
A:Title: Genetic characteristics of hepatitis delta antigens from different isolation o
A:Reference number: PD0009
A:Accession: PD0009
A:Molecule type: mRNA
A:Residues: 1-214 <TAN>
C:Superfamily: hepatitis delta virus large antigen
C:Keywords: core protein; lipoprotein; methylated carboxyl end; prenylated cysteine
P:211/Binding site: farnesyl (Cys) (covalent) #status predicted
P:211/Modified site: methyl ester carboxyl end (Cys) (in mature form) #status predicted

Query Match      1.3%; Score 7; DB 2; Length 214;
Best Local Similarity 100.0%; Pred. No. 91;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 404 KKIKKLE 410
    |||||
```

```

Db      39 KKIKKLE 45

RESULT 68
JC1062
delta large antigen - hepatitis delta virus (strain Sichuan)
N;Alternate names: HDAG
C;Species: hepatitis delta virus
C;Date: 08-Aug-1995 #sequence_revision 19-Oct-1995 #text_change 01-Aug-1997
C;Accession: JC1062
R;Liu, S.L.; Yi, T.J.; Cong, X.; Zhan, M.Y.
Chinese J. Virol. 9, 15-22, 1993
A;Title: Molecular cloning and sequencing of a hepatitis delta antigen-coding cDNA from
A;Reference number: JC1062
A;Accession: JC1062
A;Molecule type: mRNA
A;Residues: 1-214 <LIU>
C;Superfamily: hepatitis delta virus large antigen
C;Keywords: core protein; lipoprotein; methylated carboxyl end; prenylated cysteine
F;211/Binding site: farnesyl (Cys) (covalent) #status predicted
F;211/Modified site: methyl ester carboxyl end (Cys) (in mature form) #status predicted

Query Match      1.3%; Score 7; DB 2; Length 214;
Best Local Similarity 100.0%; Pred. No. 91;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy      404 KKIKKLE 410
Db      39 KKIKKLE 45

RESULT 69
S73675
hypothetical protein yjfv - Mycoplasma pneumoniae (strain ATCC 29342)
N;Alternate names: hypothetical protein P02_orf218
C;Species: Mycoplasma pneumoniae
A;Variety: ATCC 29342
C;Date: 27-Feb-1997 #sequence_revision 25-Apr-1997 #text_change 07-Dec-1999
C;Accession: S73675
R;Hammelreich, R.; Hilbert, H.; Plagens, H.; Pickl, E.; Li, B.C.; Herrmann, R.
Nucleic Acids Res. 24, 4420-4449, 1996
A;Title: Complete sequence analysis of the genome of the bacterium Mycoplasma pneumoniae
A;Reference number: S73327; MUID:97105885; PMID:8948633
A;Accession: S73675
A;Status: preliminary; nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 1-218 <HLM>
A;Cross-references: EMBL:AE000033; GB:U00089; NID:G1674022; PIDN:AB95997.1; PID:G167402
A;Note: the nucleotide sequence was submitted to the EMBL Data Library, November 1996
C;Genetics:
A;Gene: yjfv
A;Genetic code: SGC3
C;Superfamily: hypothetical protein H11024

Query Match      1.3%; Score 7; DB 2; Length 218;
Best Local Similarity 100.0%; Pred. No. 92;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy      264 TLNDIQL 270
Db      175 TLNDIQL 181

RESULT 70
T05514
hypothetical protein F13M23.70 - Arabidopsis thaliana
C;Species: Arabidopsis thaliana (mouse-ear cress)
C;Date: 23-Apr-1999 #sequence_revision 23-Apr-1999 #text_change 23-Jul-1999
C;Accession: T05514
R;Bevan, M.; Wedler, H.; Wedler, E.; Wambutt, R.; Hoheisel, J.; Mewes, H.W.; Mayer, K.F.
submitted to the Protein Sequence Database, February 1999
A;Reference number: Z15419
A;Accession: T05514

A;Molecule type: DNA
A;Residues: 1-225 <BBV>
A;Cross-references: EMBL:AL035523
A;Experimental source: cultivar Columbia; BAC clone F13M23
C;Genetics:
A;Map position: 4
A;Introns: 90/2; 128/3; 152/1
A;Note: F13M23.70

Query Match      1.3%; Score 7; DB 2; Length 225;
Best Local Similarity 100.0%; Pred. No. 95;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy      291 ERLKKLI 297
Db      205 ERLKKLI 211

RESULT 71
POBP82
antiterminal protein Q - phage 82
N;Alternate names: regulatory protein Q
C;Species: phage 82
C;Date: 30-Sep-1990 #sequence_revision 30-Sep-1990 #text_change 03-Nov-2003
C;Accession: A29791; S66585
R;Golliger, J.A.; Roberts, J.W.
J. Biol. Chem. 262, 11721-11725, 1987
A;Title: Bacteriophage 82 gene Q and Q protein: sequence, overproduction, and activity
A;Reference number: A29791; MUID:87308148; PMID:3624233
A;Accession: A29791
A;Molecule type: DNA
A;Residues: 1-229 <GOL>
A;Cross-references: GB:J02803; NID:G215364; PIDN:AAA32298.1; PID:G215365
A;Note: the authors translated the codon ACT for residue 188 as Phe and GAA for residue
R;Wahdi, A.A.; Sharples, G.J.; Mandal, T.N.; Lloyd, R.G.
J. Mol. Biol. 257, 561-573, 1996
A;Title: Holliday junction resolvases encoded by homologous rusa genes in Escherichia c
A;Reference number: S66579; MUID:96196428; PMID:8648624
A;Accession: S66585
A;Molecule type: DNA
A;Residues: 1-229 <VAH>
A;Cross-references: EMBL:CAAG3332.1; PIDN:CAAG3332.1; PID:G1051118
C;Genetics:
A;Gene: Q
C;Superfamily: antitermination protein Q
C;Keywords: DNA binding; late protein; transcription regulation

Query Match      1.3%; Score 7; DB 1; Length 229;
Best Local Similarity 100.0%; Pred. No. 97;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy      291 ERLKKLI 297
Db      144 ERLKKLI 150

RESULT 72
D90831
antitermination protein [imported] - Escherichia coli (strain O157:H7, substrain RMD 0
C;Species: Escherichia coli
C;Date: 18-Jul-2001 #sequence_revision 18-Jul-2001 #text_change 03-Nov-2003
C;Accession: D90831
R;Hayashi, T.; Makino, K.; Ohnishi, M.; Kurokawa, K.; Ishii, K.; Yokoyama, K.; Han, C.G.
gasawara, N.; Yasunaga, T.; Kuhara, S.; Shiba, T.; Hattori, M.; Shinagawa, H.
DNA Res. 8, 11-22, 2001
A;Title: Complete genome sequence of enterohemorrhagic Escherichia coli O157:H7 and gen
A;Reference number: A99629; MUID:21156231; PMID:11258796
A;Accession: D90831
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-229 <HAY>
A;Cross-references: GB:BA000007; PIDN:BA035043.1; PID:G13361084; GSPDB:GN00154
A;Experimental source: strain O157:H7, substrain RMD 0509952

```

C:Genetics:

A:Gene: ECs1620

C:Superfamily: antitermination protein Q

Query Match 1.3%; Score 7; DB 2; Length 229;

Best Local Similarity 100.0%; Pred. No. 97;

Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 291 ERLKCLI 297

Db 144 ERLKCLI 150

RESULT 73

C90912

C:Species: Escherichia coli

C:Date: 18-Jul-2001 #sequence_revision 18-Jul-2001 #text_change 03-Nov-2003

C:Accession: C90912

R:Hayashi, T.; Makino, K.; Ohnishi, M.; Kurokawa, K.; Ishii, K.; Yokoyama, K.; Han, C.G.

Gasawara, N.; Yasunaga, T.; Kuhara, S.; Shiba, T.; Hattori, M.; Shinagawa, H.

DNA Res. 8, 11-22, 2001

A:Title: Complete genome sequence of enterohemorrhagic Escherichia coli O157:H7 and gen

A:Reference number: A99629; MUID:21156231; PMID:11258796

A:Accession: C90912

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-229 <HAY>

A:Cross-references: GB:BA000007; PIDN:BA035690.1; PID:g13361733; GSPDB:GN00154

A:Experimental source: strain O157:H7, substrain RIMD 0509952

C:Genetics:

A:Gene: ECs2267

C:Superfamily: antitermination protein Q

Query Match 1.3%; Score 7; DB 2; Length 229;

Best Local Similarity 100.0%; Pred. No. 97;

Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 291 ERLKCLI 297

Db 144 ERLKCLI 150

RESULT 74

F90972

C:Species: Escherichia coli

C:Date: 18-Jul-2001 #sequence_revision 18-Jul-2001 #text_change 03-Nov-2003

C:Accession: F90972

R:Hayashi, T.; Makino, K.; Ohnishi, M.; Kurokawa, K.; Ishii, K.; Yokoyama, K.; Han, C.G.

Gasawara, N.; Yasunaga, T.; Kuhara, S.; Shiba, T.; Hattori, M.; Shinagawa, H.

DNA Res. 8, 11-22, 2001

A:Title: Complete genome sequence of enterohemorrhagic Escherichia coli O157:H7 and gen

A:Reference number: A99629; MUID:21156231; PMID:11258796

A:Accession: F90972

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-229 <HAY>

A:Cross-references: GB:BA000007; PIDN:BA036173.1; PID:g13362218; GSPDB:GN00154

A:Experimental source: strain O157:H7, substrain RIMD 0509952

C:Genetics:

A:Gene: ECs2750

C:Superfamily: antitermination protein Q

Query Match 1.3%; Score 7; DB 2; Length 229;

Best Local Similarity 100.0%; Pred. No. 97;

Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 291 ERLKCLI 297

Db 144 ERLKCLI 150

RESULT 75

G85688

C:Species: Escherichia coli

C:Date: 16-Feb-2001 #sequence_revision 16-Feb-2001 #text_change 03-Nov-2003

C:Accession: G85688

R:Perna, N.T.; Plunkett III, G.; Burland, V.; Mau, B.; Glasner, J.D.; Rose, D.J.; Mayhew

iller, L.; Grotbeck, E.J.; Davis, N.W.; Lim, A.; Dimalanta, E.; Potamousis, K.; Apodaca

Nature 409, 529-533, 2001

A:Title: Genome sequence of enterohemorrhagic Escherichia coli O157:H7.

A:Reference number: A85480; MUID:21074935; PMID:11206551

A:Accession: G85688

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-229 <SFO>

A:Cross-references: GB:AE005174; NID:g12514802; PIDN:AGS5971.1; GSPDB:GN00145; UWGP:Z1

A:Experimental source: strain O157:H7, substrain EDL933

C:Genetics:

A:Gene: Z1874

C:Superfamily: antitermination protein Q

Query Match 1.3%; Score 7; DB 2; Length 229;

Best Local Similarity 100.0%; Pred. No. 97;

Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 291 ERLKCLI 297

Db 144 ERLKCLI 150

RESULT 76

C85820

C:Species: Escherichia coli

C:Date: 16-Feb-2001 #sequence_revision 16-Feb-2001 #text_change 03-Nov-2003

C:Accession: C85820

R:Perna, N.T.; Plunkett III, G.; Burland, V.; Mau, B.; Glasner, J.D.; Rose, D.J.; Mayhew

iller, L.; Grotbeck, E.J.; Davis, N.W.; Lim, A.; Dimalanta, E.; Potamousis, K.; Apodaca

Nature 409, 529-533, 2001

A:Title: Genome sequence of enterohemorrhagic Escherichia coli O157:H7.

A:Reference number: A85480; MUID:21074935; PMID:11206551

A:Accession: C85820

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-229 <SFO>

A:Cross-references: GB:AE005174; NID:g12516127; PIDN:AGS7023.1; GSPDB:GN00145; UWGP:Z3

A:Experimental source: strain O157:H7, substrain EDL933

C:Genetics:

A:Gene: Z3114

C:Superfamily: antitermination protein Q

Query Match 1.3%; Score 7; DB 2; Length 229;

Best Local Similarity 100.0%; Pred. No. 97;

Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 291 ERLKCLI 297

Db 144 ERLKCLI 150

RESULT 77

G90292

C:Species: Sulfolobus solfataricus

C:Date: 24-May-2001 #sequence_revision 24-May-2001 #text_change 24-May-2001

C:Accession: G90292

R:She, Q.; Singh, R.K.; Confalonieri, F.; Zivanovic, V.; Allard, G.; Awayez, M.J.; Chan-

arratt, R.A.; Ragan, M.A.; Medina, N.; Peng, X.; Thi-Ngoc, H.P.; Redder, F.

submitted to GenBank, April 2001

A:Description: Sulfolobus solfataricus complete genome.

A:Reference number: A99139

A:Accession: G90292

QY 406 IKKLEKE 412
| | | | |
Db 152 IKKLEKE 158

RESULT 80
H97687
hypothetical 17.5K protein in avrA-selB intergenic region {oi57a} [imported] - Agrobact
C:Species: Agrobacterium tumefaciens
C:Date: 30-Sep-2001 #sequence_revision 30-Sep-2001 #text_change 18-Nov-2002
C:Accession: H97687
R:Goodner, B.; Hinkle, G.; Gattung, S.; Miller, N.; Blanchard, M.; Qureshi, B.; Goldman
A.; Liu, F.; Wollam, C.; Allinger, M.; Doughty, D.; Scott, C.; Lappas, C.; Markelz, B.
Science 294, 2323-2328, 2001
A:Title: Genome Sequence of the Plant Pathogen and Biotechnology Agent Agrobacterium tu
A:Reference number: A97359; MUID:21608551; PMID:11743194
A:Accession: H97687
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-237 <KUR>
A:Cross-references: GB:AB007869; PIDN:AAK88457.1; PID:gi15157960; GSPDB:GN00169
C:Genetics:
A:Gene: AGR C 4972
A:Map position: circular chromosome

Query Match 1.3%; Score 7; DB 2; Length 237;
Best Local Similarity 100.0%; Pred. No. 1e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 55 VSEELSR 61
| | | | |
Db 104 VSEELSR 110

RESULT 81
B95374
hypothetical protein Sma1644 [imported] - Sinorhizobium meliloti (strain 1021) magaplas
C:Species: Sinorhizobium meliloti
C:Date: 24-Aug-2001 #sequence_revision 24-Aug-2001 #text_change 30-Sep-2001
C:Accession: B95374
R:Barnett, M.J.; Fisher, R.F.; Jones, T.; Komp, C.; Abola, A.P.; Barloy-Hubler, F.; Bow
K.; Kalman, S.; Keating, D.H.; Palm, C.; Peck, M.C.; Surzycki, R.; Wells, D.H.; Yeh, K.C.
Proc. Natl. Acad. Sci. U.S.A. 98, 9883-9888, 2001
A:Title: Nucleotide sequence and predicted functions of the entire Sinorhizobium melilo
A:Reference number: A95262; MUID:21396509; PMID:11481432
A:Accession: B95374
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-239 <KUR>
A:Cross-references: GB:AE006469; PIDN:AAK65556.1; PID:gi14524034; GSPDB:GN00165
A:Experimental source: strain 1021, megaplasmid pSymA
R:Galibert, F.; Finau, T.M.; Long, S.R.; Puhler, A.; Abola, P.; Ampe, P.; Barloy-Hubler
Pela, D.; Chain, P.; Cowie, A.; Davis, R.W.; Dreano, S.; Federspiel, N.A.; Fisher, R.F.
L.; Hyman, R.W.; Jones, T.
Science 293, 668-672, 2001
A:Authors: Kahn, D.; Kahn, M.L.; Kalman, S.; Keating, D.H.; Kiss, B.; Komp, C.; Lelaure
heault, P.; Vandenbol, M.; Vorholter, F.J.; Weidner, S.; Wells, D.H.; Wong, K.; Yeh, K.
A:Title: The composite genome of the legume symbiont Sinorhizobium meliloti.
A:Reference number: A96039; MUID:21368234; PMID:11474104
A:Contents: annotation
C:Genetics:
A:Gene: Sma1644
A:Genome: plasmid

Query Match 1.3%; Score 7; DB 2; Length 239;
Best Local Similarity 100.0%; Pred. No. 1e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 215 HSKAVLA 221
| | | | |
Db 52 HSKAVLA 58

RESULT 82

F90885
hypothetical protein ECs2054 [imported] - Escherichia coli (strain O157:H7, substrain R1)
C:Species: Escherichia coli
C>Date: 18-Jul-2001 #sequence_revision 18-Jul-2001 #text_change 18-Jul-2001
C:Accession: F90885
R:Hayashi, T.; Makino, K.; Ohnishi, M.; Kurokawa, K.; Ishii, K.; Yokoyama, K.; Han, C.G.
gasawara, N.; Yasunaga, T.; Kuhara, S.; Shiba, T.; Hattori, M.; Shinagawa, H.
DNA Res. 8, 11-22, 2001
A:Title: Complete genome sequence of enterohemorrhagic Escherichia coli O157:H7 and genome
A:Reference number: A99829; MUID:21156231; PMID:11258796
A:Accession: F90885
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-240 <HAY>
A:Cross-references: GB:BA000007; PIDN:BA835477.1; PID:gl3361520; GSPDB:GN00154
A:Experimental source: strain O157:H7, substrain R1MD 0509952
C:Genetics:
A:Gene: ECs2054

Query Match 1.3%; Score 7; DB 2; Length 240;
Best Local Similarity 100.0%; Pred. No. 1e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 324 KLOQAQE 330
|||
Db 137 KLOQAQE 143

RESULT 83

A85733
hypothetical protein 22269 [imported] - Escherichia coli (strain O157:H7, substrain EDL933)
C:Species: Escherichia coli
C>Date: 16-Feb-2001 #sequence_revision 16-Feb-2001 #text_change 14-Sep-2001
C:Accession: A85733
R:Perna, N.T.; Plunkett III, G.; Burland, V.; Mau, B.; Glasner, J.D.; Rose, D.J.; Mayhew
Miller, L.; Grotbeck, E.J.; Davis, N.W.; Lim, A.; Dimalanta, E.; Potamousis, K.; Apodaca,
Nature 409, 529-533, 2001
A:Title: Genome sequence of enterohemorrhagic Escherichia coli O157:H7.
A:Reference number: A85480; MUID:21074935; PMID:11206551
A:Accession: A85733
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-240 <STO>
A:Cross-references: GB:AE005174; NID:gl3515249; PIDN:AA656325.1; GSPDB:GN00145; UWGP:222
A:Experimental source: strain O157:H7, substrain EDL933
C:Genetics:
A:Gene: 22269

Query Match 1.3%; Score 7; DB 2; Length 240;
Best Local Similarity 100.0%; Pred. No. 1e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 324 KLOQAQE 330
|||
Db 137 KLOQAQE 143

RESULT 84

E64897
hypothetical protein b1450 - Escherichia coli (strain K-12)
C:Species: Escherichia coli
C>Date: 12-Sep-1997 #sequence_revision 17-Sep-1997 #text_change 01-Mar-2002
C:Accession: E64897
R:Blattner, F.R.; Plunkett III, G.; Bloch, C.A.; Perna, N.T.; Burland, V.; Riley, M.; Co
A.; Rose, D.J.; Mau, B.; Shao, Y.
Science 277, 1453-1462, 1997
A:Title: The complete genome sequence of Escherichia coli K-12.
A:Reference number: A64720; MUID:97426617; PMID:9278503
A:Accession: E64897
A:Status: nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA

A:Residues: 1-240 <BLAT>
A:Cross-references: GB:AB000242; GB:U00096; NID:gl787720; PIDN:AACT4532.1; PID:gl787722
A:Experimental source: strain K-12, substrain MG1655

Query Match 1.3%; Score 7; DB 2; Length 240;
Best Local Similarity 100.0%; Pred. No. 1e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 324 KLOQAQE 330
|||
Db 137 KLOQAQE 143

RESULT 85

H72732
hypothetical protein ABE0401 - Aeropyrum pernix (strain K1)
C:Species: Aeropyrum pernix
C>Date: 20-Aug-1999 #sequence_revision 20-Aug-1999 #text_change 18-Aug-2003
C:Accession: H72732
R:Kawarabayashi, Y.; Hino, Y.; Horikawa, H.; Yamazaki, S.; Haikawa, Y.; Jin-no, K.; Taka
awa, H.; Takamiya, M.; Masuda, S.; Funahashi, T.; Tanaka, T.; Kudoh, Y.; Yamazaki, J.;
DNA Res. 6, 83-101, 1999
A:Title: Complete genome sequence of an aerobic hyper-thermophilic Crenarchaeon, Aeropy
A:Reference number: A72450; MUID:99310339; PMID:10382966
A:Accession: H72732
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-246 <KAW>
A:Cross-references: DDBJ:AP000059; NID:gs103911; PIDN:BAA79356.1; PID:gs104040
A:Experimental source: strain K1
C:Genetics:
A:Gene: ABE0401
C:Superfamily: uridylylate kinase

Query Match 1.3%; Score 7; DB 2; Length 246;
Best Local Similarity 100.0%; Pred. No. 1e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 61 ROLEDIL 67
|||
Db 186 ROLEDIL 192

RESULT 86

JH0252
myelin P0 protein precursor - human
A:Alternate names: myelin protein zero
C:Species: Homo sapiens (man)
C>Date: 31-Mar-1992 #sequence_revision 06-Dec-1996 #text_change 16-Jun-2000
R:Hayasaka, K.; Nanao, K.; Tahara, M.; Sato, W.; Takada, G.; Miura, M.; Uyemura, K.
Biochem. Biophys. Res. Commun. 180, 515-518, 1991
A:Title: Isolation and sequence determination of cDNA encoding the major structural pro
A:Reference number: JH0252; MUID:92062068; PMID:1719967
A:Accession: JH0252
A:Molecule type: mRNA
A:Residues: 1-248 <HAY>
A:Cross-references: GB:D10537; GB:D90501; NID:9220073; PIDN:BAA01395.1; PID:9220074
A:Experimental source: fetus spinal cord
R:Hayasaka, K.; Ohnishi, A.; Takada, G.; Fukushima, Y.; Murai, Y.
Biochem. Biophys. Res. Commun. 194, 1317-1322, 1993
A:Title: Mutation of the myelin P0 gene in charcot-marie-tooth neuropathy type 1.
A:Reference number: JH0704; MUID:93356807; PMID:7688964
A:Accession: JH0704
A:Molecule type: mRNA
A:Residues: 1-248 <HAZ>
R:Pham-Dinh, D.; Fourbil, Y.; Blanquet, F.; Mattei, M.G.; Roeckel, N.; Latour, P.; Chaz
Rum. Mol. Genet. 2, 2051-2054, 1993
A:Title: The major peripheral myelin protein zero gene: structure and localization in tl
A:Reference number: J39378; MUID:94154677; PMID:7509228
A:Accession: J39378
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA

```

submitted to the Protein Sequence Database, February 1999
A:Reference number: Z15385
A:Accession: T04809
A:Molecule type: DNA
A:Residues: 1-262 <35>
A:Cross-references: EMBL:AL035440
A:Experimental source: cultivar Columbia; BAC clone F10M23
C:Genetics:
A:Map position: 4
A:Introns: 74/3; 97/1; 112/3; 142/1; 216/2; 237/3
A:Note: F10M23.200
C:Superfamily: conserved hypothetical protein HI0090

Query Match 1.3%; Score 7; DB 2; Length 1
Best Local Similarity 100.0%; Pred. No. 1.1e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0

QY 21 AAPAVEA 27
|||
DB 2 AAPAVEA 8

RESULT 89
T37222
probable secreted protein - Streptomyces coelicolor
C:Species: Streptomyces coelicolor
C:Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_classification 0
A:Accession: T37222
R:Oliver, K.; Harris, D.; Parkhill, J.; Barrell, B.G.; Rajarajam, S.; et al.
submitted to the EMBL Data Library, September 1998
A:Reference number: Z21615
A:Accession: T37222
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-271 <OLL>
A:Cross-references: EMBL:AL031514; PIDN:CAA20600.1; GSPDB:O80001
A:Experimental source: strain A3(2)
C:Genetics:
A:Gene: SCOEDB:SC2H4.06c

Query Match 1.3%; Score 7; DB 2; Length 2
Best Local Similarity 100.0%; Pred. No. 1.1e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0

QY 41 GAQARTA 47
|||
DB 113 GAQARTA 119

RESULT 90
JC6142
deoxyguanosine kinase (EC 2.7.1.113) precursor - human
C:Species: Homo sapiens (man)
C:Date: 11-Apr-1997 #sequence_revision 09-May-1997 #text_classification 0
A:Accession: JC6142
R:R.Johansson, M.; Karlsson, A.
Proc. Natl. Acad. Sci. U.S.A. 93, 7258-7262, 1996
A:Title: Cloning and expression of human deoxyguanosine kinase
A:Reference number: JC6142; MID:J96293511; PMID:8692979
A:Accession: JC6142
A:Molecule type: mRNA
A:Residues: 1-277 <JOH>
A:Cross-references: GB:U41668; NID:q1477481; PIDN:AC50624
A:Note: it is uncertain whether Met-1 or Met-8 is the initiating Met
C:Comment: This enzyme is involved in mediating cytotoxicity of deoxyguanosine kinase. Thymidine kinase 2.
C:Superfamily: human deoxycytidine kinase
C:Keywords: phosphotransferase
C:Keywords: signal sequence
F:1-39/Domain: signal sequence #status predicted <SIG>
F:40-277/Product: deoxyguanosine kinase #status predicted <SIG>

Query Match 1.3%; Score 7; DB 2; Length 2
Best Local Similarity 100.0%; Pred. No. 1.2e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0

```

Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 245 QAREEE 251
 Db 205 QAREEE 211

RESULT 91
 A:Title: carboxylesterase (RC 3.1.1.1) [imported] - Yersinia pestis (strain CO92)
 C:Species: Yersinia pestis
 C>Date: 02-Nov-2001 #sequence_revision 02-Nov-2001 #text_change 27-Nov-2001
 C:Accession: A10182
 R:Parkhill, J.; Wren, B.W.; Thomson, N.R.; Titball, R.W.; Holden, M.T.G.; Prentice, M.B.; deno-Tarrage, A.M.; Chillingworth, T.; Cronin, A.; Davies, R.M.; Davis, P.; Dougan, G.; Li, M.; Rutherford, K.; Simmonds, M.; Skelton, J.; Stevens, K.; Whitehead, S.; Barrell, Nature 413, 523-527, 2001
 A:Title: Genome sequence of Yersinia pestis, the causative agent of plague.
 A:Reference number: AB0001; MUID:21470413; PMID:11586360
 A:Accession: A10182
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-280 <KUR>
 A:Cross-references: GB:AL590842; PIDN:CAC90324.1; PID:gl5979543; GSPDB:GN00175
 C:Genetics:
 A:Gene: YPO1501
 C:Superfamily: conserved hypothetical protein YJL068c
 C:Keywords: carboxylic ester hydrolase

Query Match 1.3%; Score 7; DB 2; Length 280;
 Best Local Similarity 100.0%; Pred. No. 1.2e+02;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 182 ELLEHR 188
 Db 6 ELLEHR 12

RESULT 92
 C86423
 A:Title: unknown protein, 14666-16092 [imported] - Arabidopsis thaliana
 C:Species: Arabidopsis thaliana (mouse-ear cress)
 C>Date: 02-Mar-2001 #sequence_revision 02-Mar-2001 #text_change 31-Mar-2001
 C:Accession: C86423
 R:Theologis, A.; Ecker, J.R.; Palm, C.J.; Federspiel, N.A.; Kaul, S.; White, O.; Alonso, J.; Chen, C.W.; Chung, M.K.; Conn, L.; Conway, A.B.; Dewar, T.H.; Dwyer, K.; Hansen, N.F.; Hughes, B.; Huizlar, L. Nature 408, 816-820, 2000
 A:Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim, C.; Li, J.H.; Li, Y.; Lin, X.; Liu, S.X.; Liu, Z.A.; Lueros, J.S.; Maiti, R.; Marziani, Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.
 A:Authors: Salzberg, S.L.; Schwartz, J.R.; Shinn, P.; Southwick, A.M.; Sun, H.; Tallon, M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.
 A:Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.
 A:Reference number: AB6141; MUID:21016719; PMID:11130712
 A:Accession: C86423
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-285 <STO>
 A:Cross-references: GB:AE005172; NID:gl0092456; PIDN:AGL12858.1; GSPDB:GN00141
 C:Genetics:
 A:Map position: 1

Query Match 1.3%; Score 7; DB 2; Length 285;
 Best Local Similarity 100.0%; Pred. No. 1.2e+02;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 132 SDEVGR 138
 Db 44 SDEVGR 50

RESULT 93

F64750
 A:Title: regulatory protein perR - Escherichia coli (strain K-12)
 N:Alternate names: Peroxidase resistance protein perR
 C:Species: Escherichia coli
 C>Date: 12-Sep-1997 #sequence_revision 17-Sep-1997 #text_change 01-Mar-2002
 C:Accession: F64750
 R:Blattner, F.R.; Plunkett III, G.; Bloch, C.A.; Perna, N.T.; Burland, V.; Riley, M.; C.A.; Rose, D.J.; Mau, B.; Shao, Y. Science 277, 1453-1462, 1997
 A:Title: The complete genome sequence of Escherichia coli K-12.
 A:Reference number: A64720; MUID:97426617; PMID:9278503
 A:Accession: F64750
 A:Status: nucleic acid sequence not shown; translation not shown
 A:Molecule type: DNA
 A:Residues: 1-297 <BLAT>
 A:Cross-references: GB:AE000133; GB:U00096; NID:G2367099; PIDN:AAC73357.1; PID:gl786448
 A:Experimental source: strain K-12, substrain MGL655
 C:Genetics:
 A:Gene: perR
 C:Superfamily: regulatory protein ampR
 C:Keywords: DNA binding; transcription regulation
 F:23-53/Region: regulatory protein lysR motif
 F:221-237/Domain: transmembrane #status predicted <TM>

Query Match 1.3%; Score 7; DB 2; Length 297;
 Best Local Similarity 100.0%; Pred. No. 1.2e+02;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 289 LAERLKK 295
 Db 173 LAERLKK 179

RESULT 94
 T21938
 A:Title: hypothetical protein F38A3.1 - Caenorhabditis elegans
 C:Species: Caenorhabditis elegans
 C>Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 15-Sep-2003
 C:Accession: T21938
 R:Swinburne, J. submitted to the EMBL Data Library, June 1995
 A:Reference number: Z19490
 A:Accession: T21938
 A:Status: preliminary; translated from GB/EMBL/DBJ
 A:Molecule type: DNA
 A:Residues: 1-306 <WIL>
 A:Cross-references: EMBL:Z49938; PIDN:CAA90187.1; GSPDB:GN00020; CESP:F38A3.1
 A:Experimental source: Clone F38A3
 C:Genetics:
 A:Gene: CESP:F38A3.1
 A:Map position: 2
 A:Introns: 26/2; 54/3

Query Match 1.3%; Score 7; DB 2; Length 306;
 Best Local Similarity 100.0%; Pred. No. 1.3e+02;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 77 GPGEDG 83
 Db 277 GPGEDG 283

RESULT 95
 G84701
 A:Title: hypothetical protein At2g29880 [imported] - Arabidopsis thaliana
 C:Species: Arabidopsis thaliana (mouse-ear cress)
 C>Date: 02-Feb-2001 #sequence_revision 02-Feb-2001 #text_change 02-Feb-2001
 C:Accession: G84701
 R:Lin, X.; Kaul, S.; Rounsley, S.D.; Shea, T.P.; Benito, M.I.; Town, C.D.; Fujii, C.Y.; M.; Koo, H.; Moffat, K.S.; Cronin, L.A.; Shen, M.; VanAken, S.B.; Umayam, L.; Tallon, E. euss, D.; Nierman, W.C.; White, O.; Eisen, J.A.; Salzberg, S.L.; Fraser, C.M.; Venter, C. Nature 402, 761-768, 1999
 A:Title: Sequence and analysis of chromosome 2 of the plant Arabidopsis thaliana.

A;Reference number: A84420; MUID:20083487; PMID:10617197
A;Accession: G34701
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-308 <STO>
A;Cross-references: GB:AE002093; NID:g4567213; PIDN:AAD23628.1; GSPDB:GN00139
C;Genetics:
A;Gene: At2g23980
A;Map position: 2

Query Match 1.3%; Score 7; DB 2; Length 308;
Best Local Similarity 100.0%; Pred. No. 1.3e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 336 EERHORE 342
Db 242 EERHORE 248

RESULT 96

D64303
hypothetical protein HI1305 homolog - Methanococcus jannaschii
C;Species: Methanococcus jannaschii
C;Date: 13-Sep-1996 #sequence_revision 13-Sep-1996 #text_change 07-Jul-2003
C;Accession: D64303
R;Bult, C.J.; White, O.; Olsen, G.J.; Zhou, L.; Fleischmann, R.D.; Sutton, G.G.; Blake, R.; Reich, C.I.; Overbeek, R.; Kirkness, E.F.; Weinstock, K.G.; Merrick, J.M.; Glodek, A.; Rson, J.D.; Sadow, P.W.; Hanna, M.C.; Cotton, M.D.; Roberts, K.M.; Hurst, M.A. Science 273, 1058-1073, 1996
A;Authors: Kaine, B.P.; Borodovsky, M.; Klenk, H.P.; Fraser, C.M.; Smith, H.O.; Woese, C. A;Title: Complete genome sequence of the methanogenic archaeon, Methanococcus jannaschii A;Reference number: A64300; MUID:96337999; PMID:8688087
A;Accession: D64303
A;Status: preliminary; nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 1-319 <BUL>
A;Cross-references: GS:U67461; GB:L77117; NID:GL590827; PIDN:ABE98008.1; PID:GL590830; C;Genetics:
C;Map position: REV29473-28514
C;Superfamily: thiamine monophosphate kinase

Query Match 1.3%; Score 7; DB 2; Length 319;
Best Local Similarity 100.0%; Pred. No. 1.3e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 151 GLGKEIT 157
Db 222 GLGKEIT 228

RESULT 97

T45568
hypothetical protein Fl1C1.60 - Arabidopsis thaliana
C;Species: Arabidopsis thaliana (mouse-ear cress)
C;Date: 04-Feb-2000 #sequence_revision 04-Feb-2000 #text_change 04-Feb-2000
C;Accession: T45568
R;Bargues, M.; Collado, M.C.; Navarro, P.; Terol, J.; Perez-Alonso, M.; Mewes, H.W.; May submitted to the Protein Sequence Database, December 1999
A;Reference number: 223007
A;Accession: T45568
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-322 <BAR>
A;Cross-references: EMBL:AL132976
A;Experimental source: cultivar Columbia; BAC clone Fl1C1
C;Genetics:
A;Map position: 3
A;Note: Fl1C1.60

Query Match 1.3%; Score 7; DB 2; Length 322;
Best Local Similarity 100.0%; Pred. No. 1.3e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 218 AVLARSK 224
Db 249 AVLARSK 255

RESULT 98

E64547
nitrogen fixation protein nifU homolog - Helicobacter pylori (strain 26695)
C;Species: Helicobacter pylori
C;Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 10-Sep-1999
C;Accession: E64547
R;Tomb, J.F.; White, O.; Kerlavage, A.R.; Clayton, R.A.; Sutton, G.G.; Fleischmann, R.D. Peterson, S.; Loftus, B.; Richardson, D.; Dodson, R.; Khalak, H.G.; Glodek, A.; McKenn son, J.D.; Kelley, J.M.; Cotton, M.D.; Weidman, J.M.; Fujii, C.; Bowman, C.; Watthey, L. Nature 388, 539-547, 1997
A;Authors: Wallin, E.; Hayes, W.S.; Borodovsky, M.; Karpk, P.D.; Smith, H.O.; Fraser, C. A;Title: The complete genome sequence of the gastric pathogen Helicobacter pylori. A;Reference number: A64520; MUID:97394467; PMID:9252185
A;Accession: E64547
A;Status: preliminary; nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 1-326 <TOM>

A;Cross-references: GB:AE000542; GB:AE000511; NID:G2313310; PIDN:AAD07289.1; PID:G23133
C;Superfamily: Helicobacter nitrogen fixation protein; nitrogen fixation protein homoio F:53-106/Domain: nitrogen fixation protein homology <NPH>

Query Match 1.3%; Score 7; DB 1; Length 326;
Best Local Similarity 100.0%; Pred. No. 1.3e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 465 LMKRVQD 471
Db 96 LMKRVQD 102

RESULT 99

B71960
hypothetical protein jhp0207 - Helicobacter pylori (strain J99)
C;Species: Helicobacter pylori
A;Variety: strain J99
C;Date: 12-Feb-1999 #sequence_revision 12-Feb-1999 #text_change 16-Jul-1999
C;Accession: B71960
R;Alm, R.A.; Ling, R.S.L.; Moir, D.T.; King, B.L.; Brown, E.D.; Doig, P.C.; Smith, D.R.; Ives, C.; Gibson, R.; Merberg, D.; Mills, S.D.; Jfang, Q.; Taylor, D.E.; Vovis, G.F.; Nature 387, 176-180, 1999
A;Title: Genomic sequence comparison of two unrelated isolates of the human gastric pat. A;Reference number: A71800; MUID:99120557; PMID:9923682
A;Accession: B71960
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-326 <ARN>
A;Cross-references: GB:AE001459; GB:AE001439; NID:G4154723; PIDN:AAD05790.1; PID:G41547
A;Experimental source: strain J99
C;Genetics:

A;Gene: jhp0207
C;Superfamily: Helicobacter nitrogen fixation protein; nitrogen fixation protein homolo F:53-106/Domain: nitrogen fixation protein homology <NPH>

Query Match 1.3%; Score 7; DB 2; Length 326;
Best Local Similarity 100.0%; Pred. No. 1.3e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 465 LMKRVQD 471
Db 96 LMKRVQD 102

RESULT 100

B82918
DNA-directed RNA polymerase alpha chain U0257 [imported] - Ureaplasma urealyticum
C;Species: Ureaplasma urealyticum
C;Date: 18-Aug-2000 #sequence_revision 20-Aug-2000 #text_change 02-Sep-2000
C;Accession: B82918

R;Glass, J.I.; Ieffkowitz, E.J.; Glass, J.S.; Heiner, C.R.; Chen, E.Y.; Cassell, G.H.
 submitted to GenBank, February 2000
 A:Description: The complete sequence of Ureaplasma urealyticum: Alternate views of a mir
 A:Reference number: A82870
 A:Accession: B82918
 A>Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-327 <GLA>
 A:Cross-references: GB:AE002123; GB:AF222894; MID:G6899229; PIDN:AAF30666.1; GSPDB:GN001
 A:Experimental source: serovar 3; biovar 1
 C:Genetics:
 A:Gene: rpoA; UU257
 A:Genetic code: SGC3
 C:Superfamily: DNA-directed RNA polymerase alpha chain

Query Match 1.3%; Score 7; DB 2; Length 327;
 Best Local Similarity 100.0%; Pred. No. 1.3e+02;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 248 REEEEXR 254
 |||||
 Db 254 REEEEXR 260

Search completed: June 7, 2004, 14:46:55
 Job time : 25 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: June 7, 2004, 14:38:11 ; Search time 48 Seconds
(without alignments)
3119.795 Million cell updates/sec

Title: US-10-023-529-8

Perfect score: 530

Sequence: 1 KSPGPGPRAGPEGAQRRPQ.....APSTASGQTGPQPTSARA 530

Scoring table: OLIGO
Gapop 60.0 , Gapext 60.0

Searched: 1586107 seqs, 282547505 residues

Word size : 0

Total number of hits satisfying chosen parameters: 1586107

Minimum DS seq length: 0

Maximum DS seq length: 2000000000

Post-processing: Listing first 300 summaries

Database : A Geneseq 29Jan04:*

1: Geneseqp1980s:*

2: Geneseqp1990s:*

3: Geneseqp2000s:*

4: Geneseqp2001s:*

5: Geneseqp2002s:*

6: Geneseqp2003as:*

7: Geneseqp2003bs:*

8: Geneseqp2004s:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	530	100.0	530	2	AAW49042 Human low
2	530	100.0	530	4	AAB82804 Human low
3	429	80.9	546	4	AAB82808 Human low
4	429	80.9	639	6	ABR41210 Human DIT
5	429	80.9	639	6	ABR41186 Human DIT
6	359	67.7	510	4	ABR11764 Human LDL
7	359	67.7	510	4	AAW79741 Human pro
8	356	67.2	356	6	ABO07223 Human p53
9	347	65.5	358	4	AAW78757 Human pro
10	228	43.0	1749	4	ABG00839 Novel hum
11	172	32.5	557	2	AAW49039 Rabbit lo
12	172	32.5	557	4	AAW49039 Rabbit lo
13	19	3.6	204	4	AAW83281 Human bre
14	19	3.6	386	2	AAW37883 BRCA1 mod
15	19	3.6	386	2	AAW30151 Amiro aci
16	19	3.6	528	6	ABO07222 Human p53
17	19	3.6	528	4	AAB93594 Human pro
18	17	3.2	85	5	ABP42281 Human ova
19	17	3.2	676	2	AAV17863 Neurite e
20	17	3.2	684	6	ABR43284 Human neu
21	12	2.3	495	4	AAO86301 Human pol
22	12	2.3	475	7	ADC33216 Human nov
23	9	1.7	99	4	AAU22308 Human car
24	9	1.7	99	7	ADB46276 Human car
25	9	1.7	554	4	ABB59454 Drosophila

26	9	1.7	637	6	ABU44191	Abu44191 Protein e
27	9	1.7	729	2	AAR89275	Aar89275 Yeast coa
28	8	1.5	65	2	AAV12665	AAV12665 Human 5'
29	8	1.5	156	2	AAV01195	AAV01195 Polypteti
30	8	1.5	170	3	AAO8727	AAO8727 Arabidops
31	8	1.5	174	3	AAO8726	AAO8726 Arabidops
32	8	1.5	194	4	ABG19843	ABG19843 Novel hum
33	8	1.5	224	3	AAO20570	AAO20570 Arabidops
34	8	1.5	229	5	ABU05665	ABU05665 M. tuberc
35	8	1.5	360	5	ABG77243	ABG77243 Selected
36	8	1.5	360	5	ABJ11114	ABJ11114 Yeast sel
37	8	1.5	364	5	ABJ11057	ABJ11057 Yeast sel
38	8	1.5	388	7	ADB95026	ADB95026 A. thalia
39	8	1.5	416	7	ADC31613	ADC31613 Human nov
40	8	1.5	484	4	AAW79356	AAW79356 Corynebact
41	8	1.5	484	4	AAW92676	AAW92676 C glutami
42	8	1.5	785	3	AAW41279	AAW41279 Arabidops
43	8	1.5	823	3	AAW41278	AAW41278 Arabidops
44	8	1.5	836	3	AAW41277	AAW41277 Arabidops
45	8	1.5	952	7	ADB95084	ADB95084 A. thalia
46	8	1.5	1045	4	ABG08332	ABG08332 Novel hum
47	8	1.5	1532	3	AAW40945	AAW40945 Human ORF
48	7	1.3	7	4	AAW82820	AAW82820 Human low
49	7	1.3	15	5	AAW50054	AAW50054 Prolin-e
50	7	1.3	15	7	ADD24103	ADD24103 Breast ca
51	7	1.3	18	6	ABP83398	ABP83398 G protein
52	7	1.3	25	2	AAW03444	AAW03444 HDV anti-g
53	7	1.3	27	3	AAW64949	AAW64949 Human 5'
54	7	1.3	32	2	AAW03445	AAW03445 HDV anti-g
55	7	1.3	34	6	ABP97458	ABP97458 HIV N36-b
56	7	1.3	34	6	ABP97457	ABP97457 HIV N36-b
57	7	1.3	35	2	AAW03447	AAW03447 HDV anti-g
58	7	1.3	35	2	AAW03446	AAW03446 HDV anti-g
59	7	1.3	36	4	ABG23756	ABG23756 Novel hum
60	7	1.3	36	4	ABG27966	ABG27966 Novel hum
61	7	1.3	36	4	ABG28172	ABG28172 Novel hum
62	7	1.3	37	2	AAW03443	AAW03443 HDV anti-g
63	7	1.3	37	4	ABG27959	ABG27959 Novel hum
64	7	1.3	38	2	AAW03442	AAW03442 HDV anti-g
65	7	1.3	38	4	AAW16323	AAW16323 Peptide #
66	7	1.3	38	4	ABR35315	ABR35315 Peptide #
67	7	1.3	38	4	AAW28818	AAW28818 Peptide #
68	7	1.3	38	4	ABB30145	ABB30145 Peptide #
69	7	1.3	38	4	ABB20757	ABB20757 Protein #
70	7	1.3	38	4	AAW68519	AAW68519 Human bon
71	7	1.3	38	4	AAW56146	AAW56146 Human bra
72	7	1.3	38	4	ABG50188	ABG50188 Human liv
73	7	1.3	38	4	AAW04061	AAW04061 Peptide #
74	7	1.3	38	5	ABG38100	ABG38100 Human pep
75	7	1.3	42	4	ABG28175	ABG28175 Novel hum
76	7	1.3	43	4	ABG23760	ABG23760 Novel hum
77	7	1.3	43	4	ABG23761	ABG23761 Novel hum
78	7	1.3	44	3	AAW95964	AAW95964 Construct
79	7	1.3	45	3	AAW76347	AAW76347 Fragment
80	7	1.3	45	3	ABW56241	ABW56241 Human sec
81	7	1.3	45	3	ABW56241	ABW56241 Human sec
82	7	1.3	50	2	AAW03431	AAW03431 HDV anti-g
83	7	1.3	50	2	AAW03428	AAW03428 Delta hep
84	7	1.3	50	2	AAW03429	AAW03429 HDV anti-g
85	7	1.3	50	2	AAW03438	AAW03438 HDV anti-g
86	7	1.3	50	2	AAW03439	AAW03439 HDV anti-g
87	7	1.3	50	2	AAW03437	AAW03437 HDV anti-g
88	7	1.3	50	2	AAW03434	AAW03434 HDV anti-g
89	7	1.3	50	2	AAW03436	AAW03436 HDV anti-g
90	7	1.3	50	2	AAW03440	AAW03440 HDV anti-g
91	7	1.3	51	4	ABG10316	ABG10316 Novel hum
92	7	1.3	52	4	ABG23757	ABG23757 Novel hum
93	7	1.3	52	4	AAW03441	AAW03441 HDV anti-g
94	7	1.3	57	2	ABG27957	ABG27957 Novel hum
95	7	1.3	60	7	ABG00699	ABG00699 Enterobac
96	7	1.3	70	4	ABG02022	ABG02022 Novel hum
97	7	1.3	72	4	ABG27972	ABG27972 Novel hum
98	7	1.3	73	4	AAO01466	AAO01466 Human pol

99	7	1.3	75	4	ABG25987	Abg25987 Novel hum	172	7	1.3	156	4	ABG08702	Abg08702 Novel hum
100	7	1.3	75	4	ABG28171	Abg28171 Novel hum	173	7	1.3	156	5	AAU82951	AAU82951 Human hom
101	7	1.3	75	4	ABG27974	Abg27974 Novel hum	174	7	1.3	158	4	ABG08703	Abg08703 Novel hum
102	7	1.3	76	4	ABG25983	Abg25983 Novel hum	175	7	1.3	159	6	ABU06683	ABU06683 Human mat
103	7	1.3	76	4	ABG16198	Abg16198 Novel hum	176	7	1.3	169	6	ABU70381	ABU70381 Human adi
104	7	1.3	76	5	ABP00042	ABP00042 Human ORF	177	7	1.3	173	2	AYU07084	AYU07084 Renal can
105	7	1.3	78	4	AAU59856	AAU59856 Propionib	178	7	1.3	173	4	ABG08700	ABG08700 Novel hum
106	7	1.3	78	6	ABM58375	ABM58375 Propionib	179	7	1.3	174	4	ABG02018	ABG02018 Novel hum
107	7	1.3	80	4	AAU48520	AAU48520 Propionib	180	7	1.3	174	4	ABG30142	ABG30142 Novel hum
108	7	1.3	80	6	ABM43039	ABM43039 Propionib	181	7	1.3	174	4	ABG20594	ABG20594 Novel hum
109	7	1.3	82	4	AAU80866	AAU80866 Human hae	182	7	1.3	176	2	AAU25609	AAU25609 Parietari
110	7	1.3	85	3	AAU01980	AAU01980 Human sec	183	7	1.3	176	2	ADC34859	ADC34859 Parietari
111	7	1.3	85	4	ABG27963	Abg27963 Novel hum	184	7	1.3	177	5	ABP66224	ABP66224 Bifidobac
112	7	1.3	85	4	ABG09671	Abg09671 Novel hum	185	7	1.3	186	4	ABU12097	ABU12097 Human neu
113	7	1.3	85	4	ABG28298	Abg28298 Novel hum	186	7	1.3	187	6	ADA35994	ADA35994 Acinetoba
114	7	1.3	86	5	ABP98832	ABP98832 Human pro	187	7	1.3	188	4	ABG11317	ABG11317 Human AFP
115	7	1.3	87	4	ABG23765	Abg23765 Novel hum	188	7	1.3	190	4	ABG16195	ABG16195 Novel hum
116	7	1.3	91	4	AAU81329	AAU81329 Human hae	189	7	1.3	193	3	AYU75683	AYU75683 Neisseria
117	7	1.3	91	4	AAU80872	AAU80872 Human hae	190	7	1.3	193	3	AYU75682	AYU75682 Neisseria
118	7	1.3	91	4	ABG09698	Abg09698 Novel hum	191	7	1.3	193	4	ABG27973	Abg27973 Novel hum
119	7	1.3	91	4	ABG24335	Abg24335 Protein #	192	7	1.3	194	2	AAU22383	AAU22383 Antigen t
120	7	1.3	95	4	ABG54726	Abg54726 Human liv	193	7	1.3	194	2	AAU22383	AAU22383 Antigen t
121	7	1.3	95	4	ABG09669	Abg09669 Novel hum	194	7	1.3	197	3	AAU79205	AAU79205 Soluble h
122	7	1.3	95	5	ABG42854	ABG42854 Human pep	195	7	1.3	197	6	ABU60682	ABU60682 Human sol
123	7	1.3	98	4	ABG10187	Abg10187 Novel hum	196	7	1.3	201	4	ABG03304	ABG03304 Novel hum
124	7	1.3	98	4	ABG22793	Abg22793 Novel hum	197	7	1.3	201	4	ABG27969	Abg27969 Novel hum
125	7	1.3	98	4	ABG10185	Abg10185 Novel hum	198	7	1.3	201	4	ABG23585	Abg23585 Novel hum
126	7	1.3	98	4	ABG00647	Abg00647 Novel hum	199	7	1.3	201	4	ABG23762	Abg23762 Novel hum
127	7	1.3	101	4	ABG28176	Abg28176 Novel hum	200	7	1.3	204	4	AAU38884	AAU38884 Human pol
128	7	1.3	102	4	ABG27960	Abg27960 Novel hum	201	7	1.3	205	4	AAU38883	AAU38883 Human pol
129	7	1.3	102	4	ABG27626	Abg27626 Novel hum	202	7	1.3	207	7	ADB64453	ADB64453 Human pro
130	7	1.3	104	4	ABG01408	Abg01408 Novel hum	203	7	1.3	208	4	ABG23759	Abg23759 Novel hum
131	7	1.3	105	7	ADC88181	ADC88181 Ribosomal	204	7	1.3	208	4	ABG12942	Abg12942 Novel hum
132	7	1.3	106	5	ABG43161	ABG43161 Human ova	205	7	1.3	208	4	ABG23944	Abg23944 Novel hum
133	7	1.3	114	4	ABG01409	Abg01409 Novel hum	206	7	1.3	211	5	ABP40330	ABP40330 Staphyloc
134	7	1.3	115	4	ABG27958	Abg27958 Novel hum	207	7	1.3	213	3	ABG53748	ABG53748 Arabidops
135	7	1.3	115	5	ABG70160	Abg70160 Human pre	208	7	1.3	213	3	ABG53773	ABG53773 Arabidops
136	7	1.3	117	2	AAW55515	AAW55515 H. pylori	209	7	1.3	213	3	ABG16923	ABG16923 Arabidops
137	7	1.3	117	4	AAU31353	AAU31353 Novel hum	210	7	1.3	213	3	ABG25641	ABG25641 Arabidops
138	7	1.3	118	4	ABG00818	Abg00818 Novel hum	211	7	1.3	213	3	ABG01404	Abg01404 Novel hum
139	7	1.3	118	4	ABG18560	Abg18560 Novel hum	212	7	1.3	213	4	ABG03066	Abg03066 Novel hum
140	7	1.3	118	4	ABG27964	Abg27964 Novel hum	213	7	1.3	215	3	AYU76351	AYU76351 Fragment
141	7	1.3	118	4	ABG00452	Abg00452 Novel hum	214	7	1.3	215	3	ABU56245	ABU56245 Human sec
142	7	1.3	118	4	ABG27627	Abg27627 Novel hum	215	7	1.3	215	6	ABU70766	ABU70766 Human adi
143	7	1.3	122	2	AAU01194	AAU01194 Polypepti	216	7	1.3	215	7	ADE11991	ADE11991 Human sec
144	7	1.3	127	4	AAO02070	AAO02070 Human pol	217	7	1.3	216	6	ABU70648	ABU70648 Human adi
145	7	1.3	127	7	ADE61282	ADE61282 Rat Prote	218	7	1.3	218	3	ABG53772	ABG53772 Arabidops
146	7	1.3	127	7	ADE61286	ADE61286 Rat Prote	219	7	1.3	218	3	ABG25640	ABG25640 Arabidops
147	7	1.3	129	4	ABG08365	Abg08365 Novel hum	220	7	1.3	218	3	ABG16922	ABG16922 Arabidops
148	7	1.3	130	3	ABG24902	Abg24902 Arabidops	221	7	1.3	218	3	ABG53747	ABG53747 Arabidops
149	7	1.3	133	2	AAW55583	AAW55583 H. pylori	222	7	1.3	219	3	ABG18069	ABG18069 Arabidops
150	7	1.3	133	2	AAW55697	AAW55697 H. pylori	223	7	1.3	220	4	ABG36582	ABG36582 Human FLE
151	7	1.3	133	2	AAW55584	AAW55584 H. pylori	224	7	1.3	220	5	AAO21540	AAO21540 Human rib
152	7	1.3	133	2	AAU25607	AAU25607 Parietari	225	7	1.3	225	4	ABG26532	ABG26532 Novel hum
153	7	1.3	133	7	ADC34857	ADC34857 Parietari	226	7	1.3	228	4	ABG10401	Abg10401 Novel hum
154	7	1.3	137	2	AAW55385	AAW55385 H. pylori	227	7	1.3	228	6	ABU20130	ABU20130 Protein e
155	7	1.3	138	2	ABU25610	ABU25610 Parietari	228	7	1.3	229	7	ADC00597	ADC00597 Enteroha
156	7	1.3	138	7	ADC34860	ADC34860 Parietari	229	7	1.3	229	7	ADC00070	ADC00070 Enteroha
157	7	1.3	139	5	AAU79009	AAU79009 P. juda	230	7	1.3	229	7	ADC00747	ADC00747 Enteroha
158	7	1.3	139	5	AAU79012	AAU79012 P. juda	231	7	1.3	234	4	ABG26550	ABG26550 Novel hum
159	7	1.3	139	5	AAU79008	AAU79008 P. juda	232	7	1.3	234	4	ABG27965	ABG27965 Novel hum
160	7	1.3	139	5	AAU79010	AAU79010 P. juda	233	7	1.3	234	4	ABG22794	ABG22794 Novel hum
161	7	1.3	139	5	AAU79011	AAU79011 P. juda	234	7	1.3	235	3	AAU24748	AAU24748 Plant SDF
162	7	1.3	140	7	ADE31106	ADE31106 Human dia	235	7	1.3	238	6	ABU21072	ABU21072 Protein e
163	7	1.3	142	5	ABG50085	ABG50085 Human DIT	236	7	1.3	240	6	ABU15384	ABU15384 Protein e
164	7	1.3	143	2	AAU25611	AAU25611 Parietari	237	7	1.3	240	6	ABR62263	ABR62263 GFBP-inte
165	7	1.3	143	7	ADC34861	ADC34861 Parietari	238	7	1.3	240	6	ABR62247	ABR62247 GFBP-inte
166	7	1.3	146	3	AAU24901	AAU24901 Arabidops	239	7	1.3	241	4	ABG10855	ABG10855 Novel hum
167	7	1.3	146	4	ABG20929	Abg20929 Novel hum	240	7	1.3	244	4	AAU40669	AAU40669 Human pol
168	7	1.3	150	4	ABG30011	Abg30011 Novel hum	241	7	1.3	244	4	AAU40670	AAU40670 Human pol
169	7	1.3	153	4	ABU25410	ABU25410 Novel hum	242	7	1.3	245	4	ABG03146	Abg03146 Novel hum
170	7	1.3	155	6	ABU02638	ABU02638 S. pneumo	243	7	1.3	251	4	ABG08208	ABG08208 Novel hum
171	7	1.3	156	4	ABG01403	Abg01403 Novel hum	244	7	1.3	251	7	ADB31909	ADB31909 Plant (A.

245 1.3 258 5 AAM52696 Human mye
 246 1.3 261 2 AAW70493 Human dis
 247 1.3 262 4 ABG19805 Novel hum
 248 1.3 261 4 ABG19805 Novel hum
 249 1.3 262 4 ABG19805 Novel hum
 250 1.3 263 6 ABG19805 Novel hum
 251 1.3 264 6 ABG19805 Novel hum
 252 1.3 265 6 ABG19805 Novel hum
 253 1.3 266 6 ABG19805 Novel hum
 254 1.3 267 6 ABG19805 Novel hum
 255 1.3 268 6 ABG19805 Novel hum
 256 1.3 269 6 ABG19805 Novel hum
 257 1.3 270 6 ABG19805 Novel hum
 258 1.3 271 6 ABG19805 Novel hum
 259 1.3 272 6 ABG19805 Novel hum
 260 1.3 273 6 ABG19805 Novel hum
 261 1.3 274 6 ABG19805 Novel hum
 262 1.3 275 6 ABG19805 Novel hum
 263 1.3 276 6 ABG19805 Novel hum
 264 1.3 277 6 ABG19805 Novel hum
 265 1.3 278 6 ABG19805 Novel hum
 266 1.3 279 6 ABG19805 Novel hum
 267 1.3 280 6 ABG19805 Novel hum
 268 1.3 281 6 ABG19805 Novel hum
 269 1.3 282 6 ABG19805 Novel hum
 270 1.3 283 6 ABG19805 Novel hum
 271 1.3 284 6 ABG19805 Novel hum
 272 1.3 285 6 ABG19805 Novel hum
 273 1.3 286 6 ABG19805 Novel hum
 274 1.3 287 6 ABG19805 Novel hum
 275 1.3 288 6 ABG19805 Novel hum
 276 1.3 289 6 ABG19805 Novel hum
 277 1.3 290 6 ABG19805 Novel hum
 278 1.3 291 6 ABG19805 Novel hum
 279 1.3 292 6 ABG19805 Novel hum
 280 1.3 293 6 ABG19805 Novel hum
 281 1.3 294 6 ABG19805 Novel hum
 282 1.3 295 6 ABG19805 Novel hum
 283 1.3 296 6 ABG19805 Novel hum
 284 1.3 297 6 ABG19805 Novel hum
 285 1.3 298 6 ABG19805 Novel hum
 286 1.3 299 6 ABG19805 Novel hum
 287 1.3 300 6 ABG19805 Novel hum
 288 1.3 301 6 ABG19805 Novel hum
 289 1.3 302 6 ABG19805 Novel hum
 290 1.3 303 6 ABG19805 Novel hum
 291 1.3 304 6 ABG19805 Novel hum
 292 1.3 305 6 ABG19805 Novel hum
 293 1.3 306 6 ABG19805 Novel hum
 294 1.3 307 6 ABG19805 Novel hum
 295 1.3 308 6 ABG19805 Novel hum
 296 1.3 309 6 ABG19805 Novel hum
 297 1.3 310 6 ABG19805 Novel hum
 298 1.3 311 6 ABG19805 Novel hum
 299 1.3 312 6 ABG19805 Novel hum
 300 1.3 313 6 ABG19805 Novel hum

ALIGNMENTS

RESULT 1
 AAW49042
 ID AAW49042 standard; protein; 530 AA.
 XX
 AC AAW49042;
 XX
 DT 09-NOV-1998 (first entry)
 XX Human low density lipoprotein binding protein LBP-3.
 DE
 XX Low density lipoprotein binding protein; LDL binding protein 3; LBP-3;
 XX receptor; human; atherosclerosis; diagnosis; therapy; vaccine.
 KW

XX Homo sapiens.
 OS
 XX Key Location/Qualifiers
 XX Peptide 53.59
 FT /note= "Claim 2"
 FT
 XX WO9823282-A1.
 PN
 XX 04-JUN-1998.
 PD
 XX 26-NOV-1997; 97WO-US021857.
 PF
 XX 27-NOV-1996; 96US-0031930P.
 PR
 XX 03-JUN-1997; 97US-0048547P.
 XX {BOST-} BOSTON HEART FOUND INC.
 PA
 XX Lees AM, Lees RS, Law SW, Arjona AA;
 PI
 XX WPI; 1998-322455/28.
 DR
 XX N-P8DB; AAV32839.
 DR
 XX Nucleic acid encoding low density lipoprotein binding proteins and
 PT related vectors - transformed cells, proteins, and modulators of binding,
 PT useful for treatment and diagnosis of atherosclerosis and for identifying
 PT subjects at risk.
 XX
 XX Claim 1; Fig 8; 47pp; English.
 PS
 XX This polypeptide comprises novel human low density lipoprotein (LDL)
 CC binding protein LBP-3 that is capable of binding both native and methyl
 CC LDL. Its amino acid sequence was deduced from an isolated cDNA clone (see
 CC AAV32839). cDNA clones (see AAV32834-39) and encoded rabbit and human
 CC LBPs (see AAW49037-42) are claimed. An abnormality in an aspect of LBP
 CC metabolism or structure is diagnosed of a risk for atherosclerosis. The
 CC invention provides methods for determining if an animal is at risk for
 CC atherosclerosis (e.g. for prenatal screening); methods for treating
 CC atherosclerosis (including gene therapy) using e.g. LBP polypeptides to
 CC bind LDL and thereby prevent formation of atherosclerotic plaque; and
 CC methods for treating a cell having an abnormality in LBP structure or
 CC metabolism. Pharmaceutical and vaccine compositions are also provided, as
 CC well as recombinant vectors and host cells used to produce recombinant
 CC LBP
 XX
 XX Sequence 530 AA;
 SQ
 Query Match 100.0%; Score 530; DB 2; Length 530;
 Best Local Similarity 100.0%; Pred. No. 0;
 Matches 530; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 KSSPQPEAGPEGAGQERPSQAAPAVEARGPSSQAAPRKPGEQAQARTAGSGLRDVSELS 60
 DB 1 KSSPQPEAGPEGAGQERPSQAAPAVEARGPSSQAAPRKPGEQAQARTAGSGLRDVSELS 60
 QY 61 RQLEDILSTYCVNNQGGEDGAGQEPEDAEKSTYVARNGEPEPTPVYGEKPS 120
 DB 61 RQLEDILSTYCVNNQGGEDGAGQEPEDAEKSTYVARNGEPEPTPVYGEKPS 120
 QY 121 KGPDPNTERIRQSDVGRDHRPQEKKAGLGKXETLLMOTLNTLSTPEKLAALCKKY 180
 DB 121 KGPDPNTERIRQSDVGRDHRPQEKKAGLGKXETLLMOTLNTLSTPEKLAALCKKY 180
 QY 181 AELLPEHNSOKMKLLQKKOSQOLVQEKDHLGRHSKAVLARSKLESICRELQNRSLK 240
 DB 181 AELLPEHNSOKMKLLQKKOSQOLVQEKDHLGRHSKAVLARSKLESICRELQNRSLK 240
 QY 241 BEGVQRAEEKEKKEVTSHPQVTLNDIQLQMEQHNRNSKLRQENMELAKELKLIQY 300
 DB 241 BEGVQRAEEKEKKEVTSHPQVTLNDIQLQMEQHNRNSKLRQENMELAKELKLIQY 300
 QY 301 ELREHIDKVPKHQLOOQLOVDKLOQAOEMKAEERHOREKDFLLKEAVESQRMCELM 360
 DB 301 ELREHIDKVPKHQLOOQLOVDKLOQAOEMKAEERHOREKDFLLKEAVESQRMCELM 360

Db 301 ELREEHIDKVPKHKDLQQLVDAKLQQAQEWLKEAERHOREKDFLLKEAVESQRMCELM 360
QY 361 KOETHLKOQLALYTEKFEFQNTLSKSSVFTTFQOEKEMTKKIKKLEKETTMYRSRW 420
Db 361 KOETHLKOQLALYTEKFEFQNTLSKSSVFTTFQOEKEMTKKIKKLEKETTMYRSRW 420
QY 421 ESSNKALLEMAEKTVDKLEGLQVKIORLEKLCRALQTERNDLNKRVDQLSAGQGSLL 480
Db 421 ESSNKALLEMAEKTVDKLEGLQVKIORLEKLCRALQTERNDLNKRVDQLSAGQGSLL 480
QY 481 TDGSPRRPDPGQAQAPSSPRVTEAPCYPGAPSTASGQTGPQPTTSARA 530
Db 481 TDGSPRRPDPGQAQAPSSPRVTEAPCYPGAPSTASGQTGPQPTTSARA 530
RESULT 2
AAB82804
ID AAB82804 standard; protein; 530 AA.
XX AC AAB82804;
XX DT 12-NOV-2001 (first entry)
XX DE Human low density lipoprotein binding protein 3 (LBP-3).
XX KW Low density lipoprotein binding protein 3; LBP-3; LDL; human;
XX KM atherosclerosis; antiarteriosclerotic; therapy; diagnosis; vaccine.
XX OS Homo sapiens.
XX PN W0200164874-A2.
XX PD 07-SEP-2001.
XX PF 28-FEB-2001; 2001WO-US006356.
XX PR 02-MAR-2000; 2000US-00517849.
XX PS (BOST-) BOSTON HEART FOUND INC.
XX PI Lees AM, Lees RS, Law SW, Arjona AA;
XX WPI; 2001-565505/63.
XX N-PSDB; AAH26496.
XX PT New isolated low density lipoprotein binding polypeptide for treating,
XX PT diagnosing and/or identifying therapeutic agents for atherosclerosis.
XX PS Claim 13(h); Fig 8B; 143pp; English.
XX CC The present sequence is that of the N-terminal portion of novel human low
CC density lipoprotein binding protein 3 (LBP-3). The amino acid sequence is
CC deduced from an isolated partial cDNA clone (see AAH26494). A full-length
CC sequence is given in AAB82804. Human LBP-3 is an example of claimed LBP
CC polypeptides of the invention that are capable of binding to native and
CC methylated low density lipoproteins. Also claimed are biologically active
CC fragments and analogues of LBPs, polynucleotides encoding LBPs, as well
CC as expression vectors, cells, and methods of producing the LBPs. Methods
CC of determining if an animal is at risk for atherosclerosis, methods for
CC evaluating an agent for use in treating atherosclerosis, and methods for
CC treating a cell having an abnormality in structure or metabolism of LBP
CC are claimed. Pharmaceutical compositions comprising an LBP polypeptide or
CC nucleic acid, and vaccine compositions, are also claimed
XX
XX Sequence 530 AA;
Query Match 100.0%; Score 530; DB 4; Length 530;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 530; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 KSSPGQAPGGAQERPSQAAPAVEAGPGSQAPRKPEGAQARTAQSGALRDVSEELS 60
|||||

Db 1 KSSPGQAPGGAQERPSQAAPAVEAGPGSQAPRKPEGAQARTAQSGALRDVSEELS 60
QY 61 ROLEDILSTYCVDNNOGSGEDGAQGBPAEDEDAAKSTYVARNGEPPEPTPVVYGEKPS 120
Db 61 ROLEDILSTYCVDNNOGSGEDGAQGBPAEDEDAAKSTYVARNGEPPEPTPVVYGEKPS 120
QY 121 KGDPTTEBIROSDEYGDHRDPQEKKAKGLGKEITLLMOTLNTLSTPEEKLAALCKKY 180
Db 121 KGDPTTEBIROSDEYGDHRDPQEKKAKGLGKEITLLMOTLNTLSTPEEKLAALCKKY 180
QY 181 ABLLBEHNSQVKMLLOKQSQOLVQEKDHLRGEHSAVLARSKLESICRELORNRSLK 240
Db 181 ABLLBEHNSQVKMLLOKQSQOLVQEKDHLRGEHSAVLARSKLESICRELORNRSLK 240
QY 241 EBGVQARAESEKKEVTSHFQVTLNDIQLQWEOHNRNSKLRQENMELABELKKLIEQY 300
Db 241 EBGVQARAESEKKEVTSHFQVTLNDIQLQWEOHNRNSKLRQENMELABELKKLIEQY 300
QY 301 ELREEHIDKVPKHKDLQQLVDAKLQQAQEWLKEAERHOREKDFLLKEAVESQRMCELM 360
Db 301 ELREEHIDKVPKHKDLQQLVDAKLQQAQEWLKEAERHOREKDFLLKEAVESQRMCELM 360
QY 361 KOETHLKOQLALYTEKFEFQNTLSKSSVFTTFQOEKEMTKKIKKLEKETTMYRSRW 420
Db 361 KOETHLKOQLALYTEKFEFQNTLSKSSVFTTFQOEKEMTKKIKKLEKETTMYRSRW 420
QY 421 ESSNKALLEMAEKTVDKLEGLQVKIORLEKLCRALQTERNDLNKRVDQLSAGQGSLL 480
Db 421 ESSNKALLEMAEKTVDKLEGLQVKIORLEKLCRALQTERNDLNKRVDQLSAGQGSLL 480
QY 481 TDGSPRRPDPGQAQAPSSPRVTEAPCYPGAPSTASGQTGPQPTTSARA 530
Db 481 TDGSPRRPDPGQAQAPSSPRVTEAPCYPGAPSTASGQTGPQPTTSARA 530
RESULT 3
AAB82808
ID AAB82808 standard; protein; 546 AA.
XX AC AAB82808;
XX DT 12-NOV-2001 (first entry)
XX DE Human low density lipoprotein binding protein 3 (LBP-3).
XX KW Low density lipoprotein binding protein 3; LBP-3; LDL; human;
XX KM atherosclerosis; antiarteriosclerotic; therapy; diagnosis; vaccine.
XX OS Homo sapiens.
XX PN W0200164874-A2.
XX PD 07-SEP-2001.
XX PF 28-FEB-2001; 2001WO-US006356.
XX PR 02-MAR-2000; 2000US-00517849.
XX PS (BOST-) BOSTON HEART FOUND INC.
XX PI Lees AM, Lees RS, Law SW, Arjona AA;
XX WPI; 2001-565505/63.
XX N-PSDB; AAH26501.
XX PT New isolated low density lipoprotein binding polypeptide for treating,
XX PT diagnosing and/or identifying therapeutic agents for atherosclerosis.
XX PS Claim 13(l); Fig 8A; 143pp; English.
XX CC The present sequence is that of novel human low density lipoprotein
CC binding protein 3 (LBP-3). The amino acid sequence was deduced from the

CC coding region of isolated genomic DNA (see AAH26501). It differs from the
CC sequence predicted from an LBP-3 cDNA clone (see AB92804) by the
CC presence of an additional 16 amino acids at the N-terminus (the cDNA
CC clone is 5' truncated) and by having asparagine at amino acid position
CC 130 rather than tyrosine. Human LBP-2 is an example of claimed LBP
CC polypeptides of the invention that are capable of binding to native and
CC methylated low density lipoproteins. Also claimed are biologically active
CC fragments and analogues of these LBPs, polynucleotides encoding LBPs, as
CC well as expression vectors, cells and methods of producing the LBPs. A
CC polypeptide having the amino acid residues 96-110 of the present sequence
CC is claimed (see AAH2820). Methods of determining if an animal is at risk
CC for atherosclerosis, methods for evaluating an agent for use in treating
CC atherosclerosis, and methods for treating a cell having an abnormality in
CC structure or metabolism of LBP are claimed. Pharmaceutical compositions
CC comprising an LBP polypeptide or nucleic acid, and vaccine compositions,
CC are also claimed
XX
SQ Sequence 546 AA;

Query Match 80.9%; Score 429; DB 4; Length 546;
Best Local Similarity 99.8%; Pred. No. 0;
Matches 529; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 KSPGQPEAGPEGAQRPQAPAVEAEGPGSSQAPRKPDEGAQARTAGSGLRDVSESL 60
DB 17 KSPGQPEAGPEGAQRPQAPAVEAEGPGSSQAPRKPDEGAQARTAGSGLRDVSESL 76
QY 61 RQLEDILSYCVDDNNGGDEGAQCEPAEPDAEKSRVYVARNKEPPTPVVYGEKSPS 120
DB 77 RQLEDILSYCVDDNNGGDEGAQCEPAEPDAEKSRVYVARNKEPPTPVVYGEKSPS 136
QY 121 KQDPNTEETQSDVEGDRDHRPQEKKAGKGLGKITLLMOTLNTLSTPEEKLAALCKKY 180
DB 137 KQDPNTEETQSDVEGDRDHRPQEKKAGKGLGKITLLMOTLNTLSTPEEKLAALCKKY 196
QY 181 AELLEHRNSQOMKLLQKQSQOLVQEKDHLRGEHSKAVLAESKLESICRELQRNRSIK 240
DB 197 AELLEHRNSQOMKLLQKQSQOLVQEKDHLRGEHSKAVLAESKLESICRELQRNRSIK 256
QY 241 EGVQARAREEKRKEVTSHPQVTLNDIQLOMEQHNRNSKLRQENMELAEKRLKLIQY 300
DB 257 EGVQARAREEKRKEVTSHPQVTLNDIQLOMEQHNRNSKLRQENMELAEKRLKLIQY 316
QY 301 ELREEHIDKVFHKDLQOOLVDAKLOQAEMLKEAEERHOREKDFLLAEVDSQMCML 360
DB 317 ELREEHIDKVFHKDLQOOLVDAKLOQAEMLKEAEERHOREKDFLLAEVDSQMCML 376
QY 361 KQETHLQQLALYTEKFEFPQNTLSKSESVPTTPKQEMKMTKIKKLEKETTMYRSRW 420
DB 377 KQETHLQQLALYTEKFEFPQNTLSKSESVPTTPKQEMKMTKIKKLEKETTMYRSRW 436
QY 421 ESSNKALLEMAEKTVRDKLEGLQVKIQRLEKLCRALQTERNDLNRVQDLSAGQGSLL 480
DB 437 ESSNKALLEMAEKTVRDKLEGLQVKIQRLEKLCRALQTERNDLNRVQDLSAGQGSLL 496
QY 481 TDSGPERPEGAQAPSPRVTEAPCPVGPAPSTASGQTGPQEPSTARA 530
DB 497 TDSGPERPEGAQAPSPRVTEAPCPVGPAPSTASGQTGPQEPSTARA 546

RESULT 4
ABR41210
ID ABR41210 standard; protein; 639 AA.
XX AC ABR41210;
XX DT 02-JUN-2003 (first entry)
XX DE Human DITHP extracellular signalling protein.
XX KW Human; dithp; diagnostic and therapeutic polynucleotide; diagnosis;
XX KW cancer; cell proliferative disorder; autoimmune disorder;
XX KW inflammatory disorder; infection; hormonal disorder; metabolic disorder;

KW neurological disorder; gastrointestinal disorder; transport disorder;
KW connective tissue disorder; drug screening; proteome analysis;
KW gene therapy; antisense therapy; genotyping; transgenic animal; knock in;
KW disease model; toxicological testing; transcript imaging;
KW extracellular signalling.
OS Homo sapiens.
XX WO200297031-A2.
XX 05-DEC-2002.
XX 27-MAR-2002; 2002WO-US010056.
XX 28-MAR-2001; 2001US-0279619P.
XX 29-MAR-2001; 2001US-0280067P.
XX 29-MAR-2001; 2001US-0280068P.
XX 16-MAY-2001; 2001US-0291280P.
XX 17-MAY-2001; 2001US-0291829P.
XX 17-MAY-2001; 2001US-0291849P.
XX 19-JUN-2001; 2001US-0299428P.
XX 20-JUN-2001; 2001US-0299776P.
XX 20-JUN-2001; 2001US-0300001P.
XX (INCY-) INCYTE GENOMICS INC.
XX Daffo A, Jones AL, Tran AB, Dahl CR, Gietzen D, Chinn J;
XX Dufour GE, Hillman JL, Yu JY, Tuason O, Yap PE, Amehey SR;
XX Daugherty SC, Dam TC, Liu TP, Nguyen DA, Kleefeld Y, Gerstin EH;
XX Peralta CH, David MH, Lewis SA, Chen AJ, Panzer SR, Harris B;
XX Flores V, Marwaha R, Lo A, Lan RY, Urashka ME;
XX WPI; 2003-129518/12.
XX DR N-PSDB; ACC46153.
XX Novel human diagnostic and therapeutic polypeptide useful for identifying
XX test compound which specifically binds to a polypeptide encoded by human
XX diagnostic and therapeutic polynucleotide, and to induce antibodies.

Claim 27; SEQ ID NO 745; 591bp; English.

The invention relates to novel human diagnostic and therapeutic
polynucleotides designated dithp (ACC46080-ACC46749) and to their encoded
proteins (DITHP; ABR4136-ABR4182). The invention also relates to
polynucleotide sequences at least 90% identical to the dithp cDNA
sequences of the invention; recombinant vectors, host cells and
transgenic organisms comprising a dithp nucleic acid sequence; the
recombinant production of DITHP proteins; antibodies specific for DITHP
proteins; microarrays comprising dithp nucleic acid sequences; methods of
detecting dithp nucleotide and protein sequences; methods of screening
for compounds which specifically bind a DITHP protein; and methods of
assessing the toxicity of test compounds using a dithp hybridisation
probe. Dithp nucleic acid sequences and DITHP proteins may be used in the
diagnosis of a wide variety of conditions including cancer and other cell
proliferative disorders; autoimmune or inflammatory disorders; bacterial,
viral, fungal or parasitic infections; hormonal disorders; metabolic
disorders; neurological disorders; gastrointestinal disorders; transport
disorders; and connective tissue disorders. They may also be used to
screen for modulators of protein activity or gene expression. DITHP
proteins can additionally be used in analysis of the proteome of a tissue
or cell type and to induce antibodies. The dithp nucleic acids are
additionally useful in somatic or germline gene therapy of the disorders
mentioned above, as a source of antisense sequences, as a source of
probes and primers, in genotyping and identification of individuals, in
the generation of transgenic animal models of human disease or knock in
humanised animals, in toxicological testing, and in transcript imaging.
The present sequence represents a DITHP protein which has extracellular
signalling activity. Note: The sequence data for this patent did not form
part of the printed specification, but was obtained in electronic format
directly from WIPO at ftp.wipo.int/pub/published_pct_sequences

Sequence 639 AA;

QY	241	EGVQARBEERKKEVTSHFQVTLNDIQLOMEQHNERNKLRQENMELAEKLLIEQY	300
Db	350	EGVQARBEERKKEVTSHFQVTLNDIQLOMEQHNERNKLRQENMELAEKLLIEQY	409
QY	301	ELREEHIDKVFHKDLQOOLVDKLOQAQEMKEAERHOREKDFLLKEAVESQRCMLM	360
Db	410	ELREEHIDKVFHKDLQOOLVDKLOQAQEMKEAERHOREKDFLLKEAVESQRCMLM	469
QY	361	KQETHLKKQALALYTEKPEFQNTLSKSEVFTTFQEMEMKTKIKLEKETTYMRSRW	420
Db	470	KQETHLKKQALALYTEKPEFQNTLSKSEVFTTFQEMEMKTKIKLEKETTYMRSRW	529
QY	421	ESSNKALLEVAEKTVDRDLEGLQVKIQLEKLCALQTERNDLNKRVQDLSAGQGSU	480
Db	530	ESSNKALLEVAEKTVDRDLEGLQVKIQLEKLCALQTERNDLNKRVQDLSAGQGSU	589
QY	481	TDGSPRRRPGGAQAPSSPRVTEAPCYPGAPSTASGOTGPOEPTSARA	530
Db	590	TDGSPRRRPGGAQAPSSPRVTEAPCYPGAPSTASGOTGPOEPTSARA	639
RESULT 6			
ABBI1764	ID	ABBI1764 standard; peptide; 510 AA.	
AC	ABBI1764;		
XX	11-JAN-2002	(first entry)	
DT	Human LDL binding protein homologue, SRQ ID NO:2134.		
DE	Human; cytokine; cell proliferation; tissue growth; immunomodulator; growth factor;		
KW	haematopoiesis regulation; tumour; cancer; lymphoid cell disorder; arthritis;		
KW	inhibin; chemotaxis; chemokinesis; thrombolysis; oncogenesis;		
KW	proliferation; metastasis; cancer; tumour; haematopoietic disorder;		
KW	myeloid cell disorder; lymphoid cell disorder; asthma; arthritis;		
KW	chronic inflammatory condition; proliferative retinopathy;		
KW	atherosclerosis; coronary heart disease; arterial ischaemia;		
KW	bone disorder; osteoporosis; vascular growth disorder;		
KW	tissue regeneration; wound healing; infection; immune disorder;		
KW	cell culture; drug screening; gene therapy; antiinflammatory;		
KW	antisthmatic; antiarthritic; haemostatic; antiarteriosclerotic;		
KW	cytostatic; osteopathic; vasotropic; cardiant; virucide; antibacterial;		
KW	antifungal; vulnary; antiulcer.		
XX	Homo sapiens.		
OS	WO200157188-A2.		
PN	09-AUG-2001.		
XX	05-FEB-2001; 2001WO-US003800.		
XX	03-FEB-2000; 2000US-00496914.		
PR	27-APR-2000; 2000US-00560875.		
XX	(HYSE-) HYSEQ INC.		
PA	Tang YT, Liu C, Drmanac RT;		
XX	WP1; 2001-457740/49.		
DR	N-PSDB; ABA09008.		
XX	Human proteins and DNA encoding sequences useful for preventing, treating		
PT	or ameliorating a medical condition in a mammalian subject e.g. arthritis		
PT	and cancer.		
XX	Claim 20; Page 240-241; 1963pp; English.		
PS	Sequences ABB10381-ABB12330 represent 1350 novel human polypeptides, and		
CC	sequences ABA08225-ABA09574 represent nucleic acids encoding them. The		
CC	invention also relates to vectors and recombinant host cells comprising a		

CC	nucleotide of the invention, methods of producing the novel polypeptides,		
CC	antibodies against the polypeptides, methods of detecting the nucleotides		
CC	or polypeptides in a sample, and methods of identifying compounds which		
CC	bind to polypeptides of the invention. Although novel, many of the		
CC	polypeptides of the invention have homology to known proteins, thereby		
CC	giving an insight into their probable biological activities, and hence		
CC	potential therapeutic applications. The polypeptides of the invention may		
CC	have various activities, including cytokine, cell proliferation or cell		
CC	differentiation activities; stem cell growth factor activity;		
CC	haematopoiesis regulatory activity; tissue growth activity;		
CC	immunomodulatory activity; activin- or inhibin-related activities;		
CC	chemotactic or chemokinetic activities; haemostatic, thrombotic or		
CC	thrombolytic activities; receptor or ligand activities; or may be		
CC	involved in oncogenesis, cancer cell proliferation or metastasis.		
CC	Depending on their biological activities, polypeptides and nucleotides of		
CC	the invention are useful for preventing, treating or ameliorating medical		
CC	conditions, e.g., by protein or gene therapy. Such conditions include		
CC	cancers, haematopoietic disorders (e.g. myeloid or lymphoid cell		
CC	disorders), chronic inflammatory conditions (e.g., asthma or arthritis),		
CC	proliferative retinopathy, atherosclerosis, coronary heart disease,		
CC	arterial ischaemia, bone disorders (e.g. osteoporosis), and abnormal		
CC	vascular growth. Polypeptides involved with tissue regeneration and		
CC	repair (or nucleic acids encoding them) may be used to promote wound		
CC	healing (e.g., of burns, incisions and ulcers), while those with		
CC	immunomodulatory activities may be used in the treatment of viral,		
CC	bacterial and fungal infections in addition to immune disorders.		
CC	Polypeptides with growth factor activity may be used in cell cultures to		
CC	promote cell growth. For example, such polypeptides may be used to		
CC	manipulate stem cells in culture to give rise to neuroepithelial cells		
CC	that can be used to augment or replace cells damaged by illness,		
CC	autoimmune disease or accidental damage. The polypeptides and nucleotides		
CC	may also be used in the diagnosis of the above conditions, and in drug		
CC	screening techniques. The present sequence represents a novel human		
CC	polypeptide of the invention		
XX	Sequence 510 AA;		
SQ	Query Match 67.7%; Score 359; DB 4; Length 510;		
	Best Local Similarity 99.8%; Pred. No. 0;		
	Matches 459; Conservative 0; Mismatches 1; Indels 0; Gaps 0;		
QY	1	KSSRQPEAGPEGAQERPSQAPAVAECPSSQAPRPEGAQARTAGALRDVSEELS	60
Db	28	KSSRQPEAGPEGAQERPSQAPAVAECPSSQAPRPEGAQARTAGALRDVSEELS	87
QY	61	ROLEDLITSTYCVDDNQGGPGEQAGQEPAPEDAEKSRITYVARNGEPEPTPVVYGEKPS	120
Db	88	ROLEDLITSTYCVDDNQGGPGEQAGQEPAPEDAEKSRITYVARNGEPEPTPVVYGEKPS	147
QY	121	KGDPNTEIROSDEVDGDRDRRPOEKKAKGLCKEITLLMOTLNTLSTPEKLAALCKKY	180
Db	148	KGDPNTEIROSDEVDGDRDRRPOEKKAKGLCKEITLLMOTLNTLSTPEKLAALCKKY	207
QY	181	ABELLEHRNSQOMKLLQKQSQOLVQEKDHLGEGHSAVLARSKLESCLRELQHRNSLK	240
Db	208	ABELLEHRNSQOMKLLQKQSQOLVQEKDHLGEGHSAVLARSKLESCLRELQHRNSLK	267
QY	241	REGVQARBEERKKEVTSHFQVTLNDIQLOMEQHNERNKLRQENMELAEKLLIEQY	300
Db	268	REGVQARBEERKKEVTSHFQVTLNDIQLOMEQHNERNKLRQENMELAEKLLIEQY	327
QY	301	ELREEHIDKVFHKDLQOOLVDKLOQAQEMKEAERHOREKDFLLKEAVESQRCMLM	360
Db	328	ELREEHIDKVFHKDLQOOLVDKLOQAQEMKEAERHOREKDFLLKEAVESQRCMLM	387
QY	361	KQETHLKKQALALYTEKPEFQNTLSKSEVFTTFQEMEMKTKIKLEKETTYMRSRW	420
Db	388	KQETHLKKQALALYTEKPEFQNTLSKSEVFTTFQEMEMKTKIKLEKETTYMRSRW	447
QY	421	ESSNKALLEVAEKTVDRDLEGLQVKIQLEKLCALQTERNDLNKRVQDLSAGQGSU	460
Db	448	ESSNKALLEVAEKTVDRDLEGLQVKIQLEKLCALQTERNDLNKRVQDLSAGQGSU	487

CC activity, and detecting a test agent-biased activity of the assay system.
CC Also included are modulating (M2) a p53 pathway of a cell (comprising
CC contacting a cell defective in p53 function with a candidate modulator
CC that specifically binds to a HM polypeptide comprising an HM amino acid
CC sequence, where p53 function is restored), modulating (M3) a p53 pathway
CC in a mammalian cell (comprising contacting the cell with an agent that
CC specifically binds an HM polypeptide or nucleic acid) and diagnosing (M4)
CC a disease in a patient (comprising: (a) obtaining a biological sample
CC from the patient; (b) contacting the sample with a probe for HM
CC expression; (c) comparing the results with a control; and (d) determining
CC whether the comparison indicates a likelihood disease). (M1) is useful
CC for identifying modulators of the p53 pathway. A probe for HM expression
CC is useful for diagnosing breast, colon, kidney, lung and ovarian cancer,
CC in a patient, where the cancer has greater than 25 % expression level.
CC Modulators identified by (M1) are useful in a variety of diagnostic and
CC therapeutic applications, where disease or disorder prognosis is related
CC to defects in the p53 pathway, such as, angiogenesis, apoptotic or cell
CC proliferation disorders (e.g. cancer). Another two new methods (M2 and
CC M3) are useful for modulating the p53 pathway of a cell, thus restoring
CC the p53 function of the cell, so that the cell undergoes normal
CC proliferation or progression through the cell cycle. (M2) and (M3) are
CC also useful for treating defects in the p53 pathway such as angiogenic,
CC apoptotic or cell proliferation disorders. The present sequence
CC represents a human p53 pathway modifying protein
SQ Sequence 356 AA;

Query Match 67.2%; Score 356; DB 6; Length 356;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 356; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 175 ALCKYAELEEHNSQKMLQKQSLVQSKDHLRGHSAVLARSKLSLCRELOR 234
DB 1 ALCKYAELEEHNSQKMLQKQSLVQSKDHLRGHSAVLARSKLSLCRELOR 60
QY 235 HNSLKEGVQVAREEKEKVTSHFQVTLNDIQLQMEQHNRNSKLRQENNELAERLK 294
DB 61 HNSLKEGVQVAREEKEKVTSHFQVTLNDIQLQMEQHNRNSKLRQENNELAERLK 120
QY 295 KLEQYELREHIDKVFPHKDLQQLVDAKLQQAQEMLKEABERHOREKDFLLKEAVESQ 354
DB 121 KLEQYELREHIDKVFPHKDLQQLVDAKLQQAQEMLKEABERHOREKDFLLKEAVESQ 180
QY 355 RMCELMKQSTHLKQQLALYTERPEEFQNTLSKSSSEVFTTFKQEMEKMTKKIKLEKETT 414
DB 181 RMCELMKQSTHLKQQLALYTERPEEFQNTLSKSSSEVFTTFKQEMEKMTKKIKLEKETT 240
QY 415 MYSRWESSNKALLEMAEEXTVDKLEGLQVKIQRLKLCRALQTERNDLNKRVQDLSA 474
DB 241 MYSRWESSNKALLEMAEEXTVDKLEGLQVKIQRLKLCRALQTERNDLNKRVQDLSA 300
QY 475 GGQGSILDSGPERRPBGGAQAPSSPRVTEAPCYPGAPSTASGQTGPQPTTSARA 530
DB 301 GGQGSILDSGPERRPBGGAQAPSSPRVTEAPCYPGAPSTASGQTGPQPTTSARA 356

RESULT 9

AAW78757
ID AAW78757 standard; protein; 358 AA.

AC AAW78757;

XX 06-NOV-2001 (first entry)

DE Human protein SEQ ID NO 1419.

XX Human; cytokine; cell proliferation; cell differentiation; gene therapy;
KW vaccine; peptide therapy; stem cell growth factor; haematopoiesis;
KW tissue growth factor; immunomodulatory; cancer; leukaemia;
KW nervous system disorder; arthritis; inflammation.
OS Homo sapiens.

PN W0200157190-A2.

XX 09-AUG-2001.

XX 05-FEB-2001; 2001WO-US004098.

XX 03-FEB-2000; 2000US-00496914.

PR 27-APR-2000; 2000US-00560875.

PR 20-JUN-2000; 2000US-00598075.

PR 19-JUL-2000; 2000US-00620325.

PR 01-SEP-2000; 2000US-00654936.

PR 15-SEP-2000; 2000US-00663561.

PR 20-OCT-2000; 2000US-00693325.

PR 30-NOV-2000; 2000US-00728422.

XX (HYSE-) HYSEQ INC.

XX Tang YT, Liu C, Drmanac RT, Asundi V, Zhou P, Xu C, Cao Y;

XX Ma Y, Zhao QA, Wang D, Wang J, Zhang J, Ren P, Chen R, Wang ZW;

XX Xue AJ, Yang Y, Wejhrman T, Goodrich R;

XX WPI; 2001-476283/51.

DR N-PSDB; AAK51890.

XX Nucleic acids encoding polypeptides with cytokine-like activities, useful

PT in diagnosis and gene therapy.

XX Claim 20; Page 3681-3682; 6221pp; English.

XX The invention relates to polynucleotides (AAK51456-AAK53435) and the

CC encoded polypeptides (AAW78323-AAW80302) that exhibit activity relating to

CC cytokine, cell proliferation or cell differentiation or which may induce

CC production of other cytokines in other cell populations. The

CC polynucleotides and polypeptides are useful in gene therapy, vaccines or

CC peptide therapy. The polypeptides have various cytokine-like activities,

CC e.g. stem cell growth factor activity, haematopoiesis regulating

CC activity, tissue growth factor activity, immunomodulatory activity and

CC activin/inhibin activity and may be useful in the diagnosis and/or

CC treatment of cancer, leukaemia, nervous system disorders, arthritis and

CC inflammation. Note: Records for SEQ ID NO 2110 (AAK52581), 2111

CC (AAK52582) and 3666 (AAW80020) are omitted as the relevant pages from the

CC sequence listing were missing at the time of publication

XX Sequence 358 AA;

QY Query Match 65.5%; Score 347; DB 4; Length 358;

DB Best Local Similarity 100.0%; Pred. No. 0;

QY Matches 347; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

DB 184 LEEHNSQKMLQKQSLVQSKDHLRGHSAVLARSKLSLCRELQHNRLKEEG 243

DB 12 LEEHNSQKMLQKQSLVQSKDHLRGHSAVLARSKLSLCRELQHNRLKEEG 71

QY 244 VORAREEKEKVTSHFQVTLNDIQLQMEQHNRNSKLRQENNELAERLKLIQYELR 303

DB 72 VORAREEKEKVTSHFQVTLNDIQLQMEQHNRNSKLRQENNELAERLKLIQYELR 131

QY 304 EEHIDKVPFKHDKLQQLVDAKLQQAQEMLKEABERHOREKDFLLKEAVESQRMCELMKQ 363

DB 132 EEHIDKVPFKHDKLQQLVDAKLQQAQEMLKEABERHOREKDFLLKEAVESQRMCELMKQ 191

QY 364 ETHLKQQLALYTERPEEFQNTLSKSSSEVFTTFKQEMEKMTKKIKLEKETTMYRWRWESS 423

DB 192 ETHLKQQLALYTERPEEFQNTLSKSSSEVFTTFKQEMEKMTKKIKLEKETTMYRWRWESS 251

QY 424 NKALLEMAEEXTVDKLEGLQVKIQRLKLCRALQTERNDLNKRVQDLSAGGQSLTDS 483

DB 252 NKALLEMAEEXTVDKLEGLQVKIQRLKLCRALQTERNDLNKRVQDLSAGGQSLTDS 311

QY 484 GPERRPBGGAQAPSSPRVTEAPCYPGAPSTASGQTGPQPTTSARA 530

DB 312 GPERRPBGGAQAPSSPRVTEAPCYPGAPSTASGQTGPQPTTSARA 358


```
RESULT 10
ID ABG00839 standard; protein; 1749 AA.
XX
AC ABG00839;
XX
DT 13-FEB-2002 (first entry)
XX
DE Novel human diagnostic protein #830.
XX
KW Human; chromosome mapping; gene mapping; gene therapy; forensic;
KW food supplement; medical imaging; diagnostic; genetic disorder.
XX
OS Homo sapiens.
XX
PN WO200175067-A2.
XX
PD 11-OCT-2001.
XX
PF 30-MAR-2001; 2001WO-US008631.
XX
PR 31-MAR-2000; 2000US-00540217.
XX
PR 23-AUG-2000; 2000US-00649167.
XX
PA (HYSE-) HYSEQ INC.
XX
PI Dmanac RT, Liu C, Tang YT;
XX
DR WPI; 2001-639362/73.
XX
DR N-PSDB; AAS65026.
XX
XX
XX New isolated polynucleotide and encoded polypeptides, useful in
XX diagnostics, forensics, gene mapping, identification of mutations
XX responsible for genetic disorders or other traits and to assess
XX biodiversity.
XX
XX Claim 20; SEQ ID NO 31198; 103pp; English.
XX
XX The invention relates to isolated polynucleotide (I) and polypeptide (II)
XX sequences. (I) is useful as hybridisation probes, polymerase chain
XX reaction (PCR) primers, oligomers, and for chromosome and gene mapping,
XX and in recombinant production of (II). The polynucleotides are also used
XX in diagnostics as expressed sequence tags for identifying expressed
XX genes. (I) is useful in gene therapy techniques to restore normal
XX activity of (II) or to treat disease states involving (II). (II) is
XX useful for generating antibodies against it, detecting or quantitating a
XX polypeptide in tissue, as molecular weight markers and as a food
XX supplement. (II) and its binding partners are useful in medical imaging
XX of sites expressing (II). (I) and (II) are useful for treating disorders
XX involving aberrant protein expression or biological activity. The
XX polypeptide and polynucleotide sequences have applications in
XX diagnostics, forensics, gene mapping, identification of mutations
XX responsible for genetic disorders or other traits to assess biodiversity
XX and to produce other types of data and products dependent on DNA and
XX amino acid sequences. ABG0010-ABG0377 represent novel human diagnostic
XX amino acid sequences of the invention. Note: The sequence data for this
XX patent did not appear in the printed specification, but was obtained in
XX electronic format directly from WIPO at
XX ftp.wipo.int/pub/published_pct_sequences
XX
XX Sequence 1749 AA;
XX
XX Query Match 43.0%; Score 228; DB 4; Length 1749;
XX Best Local Similarity 100.0%; Pred. No. 7.4e-211;
XX Matches 228; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
XX
XX 241 BEGVQAREEREKKEVTSHPQVTLNDIQLOMEQHNRNSKLRQENNELARLKLIEQY 300
XX
XX 438 BEGVQAREEREKKEVTSHPQVTLNDIQLOMEQHNRNSKLRQENNELARLKLIEQY 497
XX
XX 301 ELREEHIDKVPKHKDLQOQLVDAKLAQAEMLKEABERHQEKDFLLKEAVESQRMCELM 360
XX
XX |||
```

```
Db 498 ELREEHIDKVPKHKDLQOQLVDAKLAQAEMLKEABERHQEKDFLLKEAVESQRMCELM 557
Qy 361 KQOETHLKQALALYTEKFEFFQNTLSKSEVFVTFKQEMEMTKTKIKLEKETTMYRSRW 420
Db 558 KQOETHLKQALALYTEKFEFFQNTLSKSEVFVTFKQEMEMTKTKIKLEKETTMYRSRW 617
Qy 421 ESNKALLEVAERKTVDRKLEGLQVKIQRLKLCALQTERNDLNKR 468
Db 618 ESNKALLEVAERKTVDRKLEGLQVKIQRLKLCALQTERNDLNKR 665

RESULT 11
ID AAW49039 standard; protein; 557 AA.
XX
AC AAW49039;
XX
DT 09-NOV-1998 (first entry)
XX
DE Rabbit low density lipoprotein binding protein LBP-3.
XX
KW Low density lipoprotein binding protein; LDL binding protein 3; LBP-3;
KW receptor; rabbit; atherosclerosis; diagnosis; therapy; vaccine.
XX
OS Oryctolagus cuniculus.
XX
XX
XX Key Location/Qualifiers
XX Peptide 96..110
XX /note= "Claim 2"
XX
XX WO9823282-A1.
XX
XX 04-JUN-1998.
XX
XX 26-NOV-1997; 97WO-US021857.
XX
XX 27-NOV-1996; 96US-0031930P.
XX 03-JUN-1997; 97US-0048547P.
XX
XX (BOST-) BOSTON HEART FOUND INC.
XX
XX Lees AM, Lees RS, Law SW, Arjona AA;
XX
XX WPI; 1998-322455/28.
XX N-PSDB; AAV32836.
XX
XX Nucleic acid encoding low density lipoprotein binding proteins and
XX related vectors - transformed cells, proteins, and modulators of binding,
XX useful for treatment and diagnosis of atherosclerosis and for identifying
XX subjects at risk.
XX
XX Claim 1; Fig 5; 47pp; English.
XX
XX This polypeptide comprises novel rabbit low density lipoprotein (LDL)
XX binding protein LBP-3 that is capable of binding both native and methyl
XX LDL. Its amino acid sequence was deduced from rabbit abdominal aorta cDNA
XX (see AAV32836). cDNA clones (see AAV32834-39) and encoded rabbit and human
XX LBPs (see AAW49037-42) are claimed. An abnormality in an aspect of LBP
XX metabolism or structure is diagnostic of a risk for atherosclerosis. The
XX invention provides methods for determining if an animal is at risk for
XX atherosclerosis (e.g. for prenatal screening); methods for treating
XX atherosclerosis (including gene therapy) using e.g. LBP polypeptides to
XX bind LDL, and thereby prevent formation of atherosclerotic plaque; and
XX methods for treating a cell having an abnormality in LBP structure or
XX metabolism. Pharmaceutical and vaccine compositions are also provided, as
XX well as recombinant vectors and host cells used to produce recombinant
XX LBP
XX
XX Sequence 557 AA;
XX
XX Query Match 32.5%; Score 172; DB 2; Length 557;
XX Best Local Similarity 100.0%; Pred. No. 4e-157;
XX Matches 172; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```



```
QY 264 TLNDIQLQWEOHNERNSKLRQENNELAERLKKLIEQYELRESHIDKVFHKDLQOOLVDA 323
DB 280 TLNDIQLQWEOHNERNSKLRQENNELAERLKKLIEQYELRESHIDKVFHKDLQOOLVDA 339
QY 324 KLQQAQOEMLKEAERHOREKDFLLKEAVESORMCELMKQOETHLKKOOLALYTEKEEFQON 383
DB 340 KLQQAQOEMLKEAERHOREKDFLLKEAVESORMCELMKQOETHLKKOOLALYTEKEEFQON 399
QY 384 TLKSSEVPTTFKQEMERKTKIKKLEKETTMYRSRWNSSNKALLEMAEKT 435
DB 400 TLKSSEVPTTFKQEMERKTKIKKLEKETTMYRSRWNSSNKALLEMAEKT 451

RESULT 12
AAB82801
ID AAB82801 standard; protein; 557 AA.
XX
AC AAB82801;
XX
DT 12-NOV-2001 (first entry)
XX
DE Rabbit low density lipoprotein binding protein 3 (LBP-3).
XX
KW Low density lipoprotein binding protein 3; LBP-3; LDL; rabbit;
KW atherosclerosis; antiarteriosclerotic; therapy; diagnosis; vaccine.
XX
OS Oryctolagus cuniculus.
XX
PN WO200164874-A2.
XX
PD 07-SEP-2001.
XX
PF 28-FEB-2001; 2001WO-US006356.
XX
PR 02-MAR-2000; 2000US-00517849.
XX
PR 14-JUL-2000; 2000US-00616289.
XX
PA (BOST-) BOSTON HEART FOUND INC.
XX
PI Lees AM, Lees RS, Law SW, Arjona AA;
XX
DR WPI; 2001-565505/63.
XX
DR N-PSDB; AAH26491.
XX
PT New isolated low density lipoprotein binding polypeptide for treating,
PT diagnosing and/or identifying therapeutic agents for atherosclerosis.
XX
PS Claim 13(e); Fig 5; 143pp; English.
XX
CC The present sequence is that of novel rabbit low density lipoprotein
CC binding protein 1 (LBP-3). The amino acid sequence is deduced from an
CC isolated cDNA clone (see AAH26491). Rabbit LBP-3 is an example of claimed
CC polypeptides of the invention, termed LBPs, that are capable of binding
CC to native and methylated low density lipoproteins. Also claimed are
CC biologically active fragments and analogues of LBPs, polynucleotides
CC encoding LBPs, as well as expression vectors, cells and methods of
CC producing the LBPs. A polypeptide having the amino acid residues 96-110
CC of the present sequence is claimed (see AAB82819). Methods of determining
CC if an animal is at risk for atherosclerosis, methods for evaluating an
CC agent for use in treating atherosclerosis, and methods for treating a
CC cell having an abnormality in structure or metabolism of LBP are claimed.
CC Pharmaceutical compositions comprising an LBP polypeptide or nucleic
CC acid, and vaccine compositions, are also claimed
XX
SQ Sequence 557 AA;

Query Match 32.5%; Score 172; DB 4; Length 557;
Best Local Similarity 100.0%; Pred. No. 4e-157;
Matches 172; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 264 TLNDIQLQWEOHNERNSKLRQENNELAERLKKLIEQYELRESHIDKVFHKDLQOOLVDA 323
DB 280 TLNDIQLQWEOHNERNSKLRQENNELAERLKKLIEQYELRESHIDKVFHKDLQOOLVDA 339
QY 324 KLQQAQOEMLKEAERHOREKDFLLKEAVESORMCELMKQOETHLKKOOLALYTEKEEFQON 383
DB 340 KLQQAQOEMLKEAERHOREKDFLLKEAVESORMCELMKQOETHLKKOOLALYTEKEEFQON 399
QY 384 TLKSSEVPTTFKQEMERKTKIKKLEKETTMYRSRWNSSNKALLEMAEKT 435
DB 400 TLKSSEVPTTFKQEMERKTKIKKLEKETTMYRSRWNSSNKALLEMAEKT 451
```

```
DB 280 TLNDIQLQWEOHNERNSKLRQENNELAERLKKLIEQYELRESHIDKVFHKDLQOOLVDA 339
QY 324 KLQQAQOEMLKEAERHOREKDFLLKEAVESORMCELMKQOETHLKKOOLALYTEKEEFQON 383
DB 340 KLQQAQOEMLKEAERHOREKDFLLKEAVESORMCELMKQOETHLKKOOLALYTEKEEFQON 399
QY 384 TLKSSEVPTTFKQEMERKTKIKKLEKETTMYRSRWNSSNKALLEMAEKT 435
DB 400 TLKSSEVPTTFKQEMERKTKIKKLEKETTMYRSRWNSSNKALLEMAEKT 451

RESULT 13
AAB63259
ID AAB63259 standard; protein; 204 AA.
XX
AC AAB63259;
XX
DT 26-MAR-2001 (first entry)
XX
DE Human breast cancer associated antigen protein sequence SEQ ID NO:621.
XX
KW Human; breast cancer; gastric cancer; prostate cancer; diagnosis;
KW cancer associated antigen; cytostatic; cancer vaccine.
XX
OS Homo sapiens.
XX
PN WO200073801-A2.
XX
PD 07-DEC-2000.
XX
PF 26-MAY-2000; 2000WO-US014749.
XX
PR 28-MAY-1999; 99US-0136526P.
XX
PR 10-SEP-1999; 99US-0153454P.
XX
PA (LUDW-) LUDWIG INST CANCER RES.
XX
PI Obata Y;
XX
DR WPI; 2001-025274/03.
XX
PT Nucleic acids encoding breast, gastric and prostate cancer associated
PT antigen precursors, useful for diagnosing and treating a condition
PT characterized by expression of an abnormal amount of a protein, e.g.
PT cancer.
XX
PS Example 1; Page 486-487; 799pp; English.
XX
CC AAF22422 to AAF22626, AAF22627 to AAF22773 and AAF22774 to AAF23014
CC represent nucleotide sequences encoding human breast, gastric and
CC prostate cancer associated antigen precursors (CAAP) respectively.
CC AAB63232 to AAB63467, AAB63468 to AAB63721 and AAB63722 to AAB63970
CC represent human breast, gastric and prostate CAAP protein sequence
CC respectively. CAAPs have cytostatic activity and can be used in the
CC production of cancer vaccines. The human CAAP proteins, peptides, nucleic
CC acids or anti-CAAP antibodies are useful for diagnosing and treating a
CC condition characterised by expression of an abnormal amount of a protein,
CC e.g. cancer
XX
SQ Sequence 204 AA;

Query Match 3.6%; Score 19; DB 4; Length 204;
Best Local Similarity 100.0%; Pred. No. 1.4e-09;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 163 LNTLSTPEEKLAALCKKYA 181
DB 1 LNTLSTPEEKLAALCKKYA 19

RESULT 14
AAB37883
ID AAB37883 standard; protein; 386 AA.
```

```

XX AAW37883;
XX AC
XX XX
XX DT 28-AUG-1998 (first entry)
XX DE BRCA1 modulator protein 091-132Q20.
XX KW BRCA1 modulator protein; 091-132Q20; breast cancer antigen 1;
XX KW tumour suppressor protein; diagnosis; therapy; human.
XX OS Homo sapiens.
XX FH Key Location/Qualifiers
XX FT Domain 124..143
XX FT /note= "leucine zipper motif"
XX PN WO9810066-A1.
XX PD 12-MAR-1998.
XX PF 06-AUG-1997; 97WO-US013944.
XX PR 04-SEP-1996; 96US-0025601P.
XX PA (ONVX-) ONVX PHARM INC.
XX PI Rubinfeld B, Polakis P, Ligenfelter C, Vuong TT;
XX WPI; 1998-193616/17.
XX DR N-PSDB; AAV29064.
XX PT Breast cancer antigen 1 modulator protein - useful for diagnosing
XX PT diseases involving unwanted cell growth, e.g. breast cancer, and for
XX PT producing therapeutics for treatment of such diseases.
XX PS Example 1; Fig 3; 73pp; English.
XX CC This polypeptide comprises a 46 kDa BRCA1 modulator protein that binds to
XX CC the tumour suppressor gene product BRC1, and which is characterised by a
XX CC leucine zipper motif. Its amino acid sequence was deduced from the
XX CC nucleotide sequence of a cDNA clone (see AAV29064), designated 091-132Q20
XX CC (ATCC 98143), isolated from a HeLa cell cDNA library using a yeast two-
XX CC hybrid assay. 3 cDNA clones (see also AAV29062 and AAV29063) coding for
XX CC BRCA1 modulator proteins (see AAW37881-83) have been characterised.
XX CC Vectors and host cells comprising the isolated nucleic acid sequences are
XX CC claimed, as well as a process for producing BRCA1 modulator protein by
XX CC culturing these host cells. BRCA1 modulator proteins and nucleic acids
XX CC can be used to diagnose diseases involving unwanted cell growth, e.g.
XX CC breast cancer, and to identify compounds that alter BRCA1 interaction
XX CC with BRCA1 modulators for the treatment of such diseases
XX SQ Sequence 386 AA;
Query Match 3.6%; Score 19; DB 2; Length 386;
Best Local Similarity 100.0%; Pred. No. 2.5e-09;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 163 LNTLSTPEKLAALCKKYA 181
DB 135 LNTLSTPEKLAALCKKYA 153
RESULT 15
AAV30151
ID AAY30151 standard; protein; 386 AA.
XX AC AAY30151;
XX XX
XX DT 27-OCT-1999 (first entry)
XX DE Amino acid sequence of a BRCA1 modulator protein.
XX KW Modulator protein; BRCA1; tumour suppressor protein; breast cancer;

```

```

KW ovarian cancer; cell growth; cell proliferation.
XX OS Homo sapiens.
XX FH Key Location/Qualifiers
XX FT Region 125..143
XX FT /note= "leucine zipper motif"
XX PN US948643-A.
XX PD 07-SEP-1999.
XX PF 13-AUG-1997; 97US-00968751.
XX PR 13-AUG-1997; 97US-00968751.
XX PA (ONVX-) ONVX PHARM INC.
XX PI Rubinfeld B, Ligenfelter C, Vuong TT, Polakis PG;
XX WPI; 1999-517952/43.
XX DR N-PSDB; AAX86756.
XX PT Modulator proteins that bind to and modulate the activity of the BRCA1
XX PT tumour suppressor gene product, useful for the treatment of ovarian and
XX PT breast cancer.
XX PS Example 1; Fig 3; 35pp; English.
XX CC The present sequence represents a modulator protein, that binds to and
XX CC modulate the activity of the BRCA1 gene product (BRCA1). The BRCA1
XX CC protein has been characterized as a tumour suppressor protein.
XX CC Alterations in the amino acid sequence of BRCA1 causes breast and ovarian
XX CC cancers by removing the controls on cell growth and proliferation.
XX CC Research has shown that different regions on the BRCA1 molecule have
XX CC different effects on cell growth and tumour suppression (e.g. full length
XX CC truncated BRCA1 has no effect on breast cancer cell growth but will
XX CC inhibit ovarian cancer cell growth). It has been suggested that different
XX CC host cell factors (e.g. proteins) interact with different regions of the
XX CC BRCA1 to control its function. The identification of these proteins (e.g.
XX CC BRCA1MP) will facilitate the development of novel diagnostic methods and
XX CC new therapeutics for identifying and treating cancers caused by changes
XX CC in the expression or activity of BRCA1
XX SQ Sequence 386 AA;
Query Match 3.6%; Score 19; DB 2; Length 386;
Best Local Similarity 100.0%; Pred. No. 2.5e-09;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 163 LNTLSTPEKLAALCKKYA 181
DB 135 LNTLSTPEKLAALCKKYA 153
RESULT 16
ABO07222
ID ABO07222 standard; protein; 395 AA.
XX AC ABO07222;
XX XX
XX DT 13-AUG-2003 (first entry)
XX DE Human p53 modifying protein, SEQ ID 182.
XX KW Human; p53 modifier; cytostatic; cancer; cytostatic; antiangiogenic;
XX KW antiapoptotic; p53 pathway; breast cancer; colon cancer; kidney cancer;
XX KW lung cancer; ovarian cancer; angiogenesis; cell cycle;
XX KW apoptotic disorder; cell proliferation disorder.
XX OS Homo sapiens.
XX PN WO200299122-A1.

```


DE Human ovarian antigen HTJA50, SEQ ID NO:3413.
XX Human; ovarian antigen; ovary; ovarian; breast; cancer; tumour;
KW ovarian cancer; breast cancer; tumour; reproductive system disorder;
KW infertility; pregnancy disorder; anovulation; polycystic ovary syndrome;
KW PCOS; ovarian cyst; dysmenorrhoea; endocrine disorder; infection;
KW inflammatory condition; immune disorder; blood disorder;
KW cardiovascular disorder; respiratory disorder; neurological disorder;
KW gastrointestinal disorder; urinary system disorder; drug screening;
KW gene therapy; chromosome mapping; forensic analysis;
KW antibody preparation; cytostatic; immunomodulatory; neuroprotective;
KW antiinflammatory; gynaecological; reproductive.
XX Homo sapiens.
OS
XX
FN WO200200677-A1.
XX
XX 03-JAN-2002.
XX
XX 07-JUN-2001; 2001WO-US018569.
XX
XX 07-JUN-2000; 2000US-0209467P.
XX
XX (HUMA-) HUMAN GENOME SCI INC.
XX
XX Birse CE, Rosen CA;
PI
XX
XX WPI; 2002-147878/19.
DR N-PSDB; ABQ55358.
XX
XX Isolated nucleic acid molecules encoding novel ovarian polypeptides,
PT useful in the prevention, treatment and diagnosis of cancer (e.g. ovarian
PT cancer), immune disorders, cardiovascular disorders and neurological
PT diseases.
XX
XX Claim 11; SEQ ID NO 3413; 2922pp; English.
XX
XX The invention relates to 2175 novel human ovarian antigens (ABP41054-
CC ABP43228) and to cDNAs encoding them (ABQ54131-ABQ56305), and also
CC encompasses polypeptides 90% identical and polynucleotides 95% identical
CC to the sequences of the invention. The invention additionally relates to
CC recombinant vectors and host cells comprising human ovarian antigen
CC polynucleotides, antibodies against human ovarian antigens, and the use
CC of ovarian antigen polynucleotides and polypeptides in diagnosing,
CC treating, prognosing or preventing various ovary and/or breast-related
CC disorders. Such conditions include ovarian cancer and breast cancer, and
CC metastatic tumours of ovarian or breast origin, reproductive system
CC disorders (e.g., infertility, disorders of pregnancy, anovulation,
CC polycystic ovary syndrome, ovarian cysts, and dysmenorrhoea), endocrine
CC disorders, infections (e.g., chlamydia, HIV, toxoplasmosis, and toxic
CC shock syndrome), inflammatory conditions (e.g., mastitis, oophoritis and
CC vaginitis), immune disorders (e.g., congenital and acquired
CC immunodeficiencies, autoimmune oophoritis, systemic lupus erythematosus),
CC blood-related disorders (e.g., anaemia), cardiovascular disorders,
CC respiratory disorders, neurological disorders, gastrointestinal disorders
CC and urinary system disorders. Ovarian antigen polypeptides and
CC polynucleotides may also be used in screening for compounds which
CC modulate ovarian antigen expression or activity. The polynucleotides may
CC further be used for gene therapy, chromosome mapping, in the
CC identification of individuals and in forensic analysis, and the
CC polypeptides may be used as food additives or to prepare antibodies
CC useful in disease diagnosis, drug targeting and phenotyping. The present
CC sequence represents a human ovarian antigen of the invention. Note: The
CC sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences
XX
XX Sequence 85 AA;
XX
XX Query Match 3.2%; Score 17; DB 5; Length 85;
XX Best Local Similarity 100.0%; Pred. No. 5.9e-08;
XX Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 397 QEMKMTKKIKKLEKET 413
Db |||||
52 QEMKMTKKIKKLEKET 68
|||
RESULT 19
AAV17863
ID AAV17863 standard; protein; 676 AA.
XX
AC AAV17863;
XX
DT 16-AUG-1999 (first entry)
XX
XX Neurite extending activity protein.
DE
XX Neurite extending activity; anti-dementia; memory; brain function;
KW dementia.
KW
XX Mus sp.
OS
XX
FN JP11147897-A.
XX
PD 02-JUN-1999.
XX
XX 13-NOV-1997; 97JP-00331242.
PF
XX 13-NOV-1997; 97JP-00331242.
PR
XX (AGEN) AGENCY OF IND SCI & TECHNOLOGY.
PA
XX
XX WPI; 1999-379889/32.
DR N-PSDB; AAX80156.
XX
XX New protein with neurite extending activity - useful for treating
PT dementia.
PT
XX
PS Claim 1; Page 5-8; 9pp; Japanese.
XX
XX The present sequence represents a protein (I) which has neurite extending
CC activity. (I) may be administered to patients to prevent dementia or to
CC improve memory and brain function. (I) is also used as a neurite-
CC extending agent
XX
XX Sequence 676 AA;
SQ
XX
XX Query Match 3.2%; Score 17; DB 2; Length 676;
XX Best Local Similarity 100.0%; Pred. No. 3.4e-07;
XX Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 220 LARSKLESICRELQREN 236
Db |||||
207 LARSKLESICRELQREN 223
|||
RESULT 20
ABR43284
ID ABR43284 standard; protein; 684 AA.
XX
AC ABR43284;
XX
XX 09-JUL-2003 (first entry)
DT
XX
DE Human neurotransmission-associated protein NTRAN-14 SEQ ID NO:14.
XX
XX Human; neurotransmission-associated protein; NTRAN; cytostatic; anti-HIV;
KW antiarteriosclerotic; anticonvulsant; nootropic; neuroprotective; stroke;
KW cerebroprotective; antiallergic; antiinflammatory; thyromimetic; cancer;
KW antidiabetic; gene therapy; cell proliferative disorder; atherosclerosis;
KW neurological disorder; epilepsy; Huntington's disease; immune disorder;
KW inflammatory disorder; AIDS; allergy; developmental disorder; diabetes;
KW hypothyroidism; Cushing's syndrome; endocrine disorder; infection.
XX
XX Homo sapiens.
OS

XX WO2003025129-A2.
XX 27-MAR-2003.
XX 12-SEP-2002; 2002WO-US029219.
XX 14-SEP-2001; 2001US-0322180P.
XX 28-SEP-2001; 2001US-0326096P.
XX 04-OCT-2001; 2001US-0327446P.
XX 26-OCT-2001; 2001US-0345837P.
XX 02-NOV-2001; 2001US-0343903P.
XX 27-NOV-2001; 2001US-034020P.
XX 07-DEC-2001; 2001US-0340226P.
XX 04-JAN-2002; 2002US-0345008P.
XX 18-MAR-2002; 2002US-0365645P.
XX 10-MAY-2002; 2002US-0379887P.
XX (INCY-) INCYTE GENOMICS INC.
XX Honchell CD, Warren BA, Borowsky ML, Griffin JA, Li JX, Lee SY;
PI Yue H, Forsythe LJ, Marquis JP, Gietzen KJ, Baughn MR, Tran UK;
PI Lehr-Mason PM, Tang YT, Ramkumar J, Emerling BM, Lee EA, Elliott VS;
PI Hafalia AJA, Duggan BM, Chawla NK, Kable AE, Chang H, Khare R;
PI Becha SD, Jin P, Lee S;
XX WPI; 2003-363137/34.
XX N-PSDB; ACC68992.
XX New human neurotransmission associated proteins (NTRAN), useful for
PT diagnosing, treating and preventing diseases or conditions associated
PT with the aberrant NTRAN expression e.g. cancer, AIDS, diabetes, epilepsy,
PT or infections.
XX Claim 1; Page 201-203; 240pp; English.
XX ACC68979 to ACC69003 encode the human neurotransmission-associated
CC proteins given in ABR43271 to ABR43295, designated NTRAN-1 to NTRAN-25
CC (1). (i) have cytostatic, antiarteriosclerotic, anticonvulsant,
CC nootropic, neuroprotective, cerebroprotective, anti-HIV, antiallergic,
CC antiinflammatory, thymometric and antidiabetic activities, and can be
CC used in gene therapy. The NTRAN polypeptides and polynucleotides are
CC useful in diagnosing, treating and preventing diseases or conditions
CC associated with the decreased expression or overexpression of NTRAN, such
CC as cell proliferative (e.g. cancer, atherosclerosis), neurological (e.g.
CC epilepsy, Huntington's disease, stroke), immune/inflammatory (e.g. AIDS,
CC allergies), developmental (e.g. hypothyroidism, Cushing's syndrome) or
CC endocrine (e.g. diabetes) disorders, or infections. They are also useful
CC in assessing the effects of exogenous compounds on the expression of
CC nucleic acid and amino acid sequences of NTRAN. The NTRANs or their
CC fragments are useful in screening compounds for effectiveness as agonist
CC or antagonist of the polypeptides, or in altering the expression of the
CC target polynucleotide and compounds that specifically bind to or modulate
CC the activity of the polypeptide
XX Sequence 684 AA;
Query Match 3.2%; Score 17; DB 6; Length 684;
Best Local Similarity 100.0%; Pred. No. 3.4e-07;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY 220 LARSKLESCLRELQRHN 236
DB 209 LARSKLESCLRELQRHN 225
RESULT 21
AAO08301
ID AAO08301 standard; protein; 475 AA.
XX AAO08301;
AC AAO08301;
XX 06-NOV-2001 (first entry)
DT

XX Human polypeptide SEQ ID NO 22193.
XX Human; cytokine; cell proliferation; cell differentiation; gene therapy;
KW vaccine; peptide therapy; stem cell growth factor; haematopoiesis;
KW tissue growth factor; immunomodulatory; cancer; leukaemia;
KW nervous system disorders; arthritis; inflammation.
XX Homo sapiens.
XX WO200164835-A2.
XX 07-SEP-2001.
XX 26-FEB-2001; 2001WO-US004927.
XX 28-FEB-2000; 2000US-00515126.
XX 18-MAY-2000; 2000US-00577409.
XX (HYSB-) HYSSEQ INC.
XX Tang YT, Liu C, Drmanac RT;
XX WPI; 2001-514838/56.
XX N-PSDB; AAI88232.
XX Isolated nucleic acids and polypeptides, useful for preventing diagnosing
PT and treating e.g. leukemia, inflammation and immune disorders.
XX Claim 20; SEQ ID NO 22193; 1399pp + Sequence Listing; English.
XX The invention relates to human polynucleotides (AAI79941-AAI93841) and
CC the encoded proteins (AAO00010-AAO13910) that exhibit activity elating to
CC cytokine, cell proliferation or cell differentiation or which may induce
CC production of other cytokines in other cell populations. The
CC polynucleotides and polypeptides are useful in gene therapy, vaccines or
CC peptide therapy. The polypeptides have various cytokine-like activities,
CC e.g. stem cell growth factor activity, haematopoiesis regulating
CC activity, tissue growth factor activity, immunomodulatory activity and
CC activin/inhibin activity and may be useful in the diagnosis and/or
CC treatment of cancer, leukaemia, nervous system disorders, arthritis and
CC inflammation. Note: The sequence data for this patent did not form part
CC of the printed specification, but was obtained in electronic format
CC directly from WIPO at ffp.wipo.int/pub/published_pct_sequences
XX Sequence 475 AA;
Query Match 2.3%; Score 12; DB 4; Length 475;
Best Local Similarity 100.0%; Pred. No. 0.017;
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY 245 QRAREEERKKE 256
DB 33 QRAREEERKKE 44
RESULT 22
ADC3216
ID ADC3216 standard; protein; 475 AA.
XX ADC3216;
AC ADC3216;
XX 18-DEC-2003 (first entry)
XX Human novel contig-encoded polypeptide sequence, SEQ ID NO:3298.
XX Human; diagnostic; drug screening; forensics; gene mapping;
KW biodiversity assessment; Parkinson's disease; Alzheimer's disease;
KW neurodegenerative diseases; anaemia; platelet disorder; wound; Burns;
KW ulcers; osteoporosis; autoimmune disease; cancer;
KW molecular weight marker; food supplement; antiparkinsonian; nootropic;
KW neuroprotective; antianaemic; anticoagulant; thrombolytic; vulnary;
KW antitumor; osteopathic; immunosuppressive; antiinflammatory; cytostatic;

KW gene therapy; chromosome 6.
 XX Homo sapiens.
 OS WO2003029271-A2.
 XX 10-APR-2003.
 XX 24-SEP-2002; 2002WO-US030474.
 XX 24-SEP-2001; 2001US-0324631P.
 XX (HYSE-) HYSEQ INC.
 XX Tang TY, Zhang J, Ren F, Xue AJ, Zhao QA, Wang J, Wehrman T;
 PI Zhou P, Chosh M, Wang D, Ma Y, Asundi V, Wang Z, Weng G;
 PI Haley-Vicente D, Drmanac RT;
 XX WPI; 2003-371981/35.
 DR N-PSDB; ABC32449.
 XX
 PT New polynucleotide and polypeptide useful for diagnosing, preventing or
 PT treating conditions such as neurodegenerative diseases, anemias, platelet
 PT disorders, wounds, burns, ulcers, osteoporosis, autoimmune diseases or
 PT cancer.
 XX
 PS Example 2; SEQ ID NO 3298; 1185pp; English.
 XX
 CC The invention relates to 971 novel human cDNA sequences (ADC29919-
 CC ADC30889) and the polypeptides they encode (ADC30890-ADC31860). The
 CC invention also relates to nucleic acid sequences over 99% identical with
 CC the novel human cDNAs. The invention additionally encompasses expression
 CC vectors and host cells comprising a nucleic acid of the invention; the
 CC recombinant production of a polypeptide of the invention; an antibody
 CC against a polypeptide of the invention; a method of detecting
 CC polynucleotides or polypeptides of the invention; and methods of
 CC identifying a compound which binds to a polypeptide of the invention. The
 CC invention further discloses methods of preventing, treating or
 CC ameliorating a medical condition; kits comprising polynucleotide probes
 CC and/or monoclonal antibodies for carrying out the methods of the
 CC invention; methods for the identification of compounds that modulate the
 CC expression or activity of the polynucleotide and/or polypeptide; and 767
 CC contig sequences corresponding to the cDNA sequences of the invention
 CC (ADC31861-ADC32627) and the polypeptides encoded by the contigs (ADC32628
 CC -ADC33394). The nucleic acids and polypeptides of the invention are
 CC useful in diagnostics, drug screening, forensics, gene mapping, in the
 CC identification of mutations responsible for genetic disorders or other
 CC traits, for assessing biodiversity, and in producing many other types of
 CC data and products dependent on DNA and amino acid sequences. They are
 CC also used for treating diseases such as Parkinson's disease, Alzheimer's
 CC disease and other neurodegenerative diseases, anaemia, platelet
 CC disorders, wounds, burns, ulcers, osteoporosis, autoimmune diseases or
 CC cancer. The nucleic acids may also be used as hybridisation probes or
 CC primers, and in the recombinant production of a protein. The polypeptides
 CC are also useful in generating antibodies, as molecular weight markers,
 CC and as food supplements. The present sequence represents a human contig-
 CC encoded polypeptide sequence used in an example of the invention. Note:
 CC The sequence data for this patent did not form part of the printed
 CC specification, but was obtained in electronic format directly from WIPO
 CC at ftp.wipo.int/pub/published_pct_sequences.
 XX
 SQ Sequence 475 AA;
 Query Match 2.3%; Score 12; DB 7; Length 475;
 Best Local Similarity 100.0%; Pred. No. 0.017;
 Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 245 QPAREEEKKE 256
 DB 33 QPAREEEKKE 44
 RESULT 23

AAU22308
 ID AAU22308 standard; protein; 99 AA.
 XX
 AC AAU22308;
 XX
 DT 18-DEC-2001 (first entry)
 XX
 DE Human cardiovascular system antigen polypeptide SEQ ID NO 1082.
 XX
 KW Cardiovascular system antigen; human; mouse; rabbit; goat; horse; cat;
 KW chicken; sheep; immunosuppressive; antiarthritic; vasotropic; dog;
 KW antirheumatic; antiproliferative; cytostatic; cardant; neuroprotective;
 KW cerebroprotective; nootropic; antibacterial; virucide; fungicide; cancer;
 KW ophthalmological; vulnery; gene therapy; autoimmune disease; neoplasm;
 KW hyperproliferative disorder; breast; liver; cardiovascular disorder;
 KW cerebrovascular disorder; nervous system disorder; bacterial infection;
 KW fungal infection; viral infection; ocular disorder; endocrine disorder;
 KW gastrointestinal disorder; renal disorder; respiratory disorder;
 KW wound healing; skin aging; organ transplantation; tissue regeneration;
 KW anti-infertility.
 XX
 OS Homo sapiens.
 XX
 PN WO200155321-A2.
 XX
 PD 02-AUG-2001.
 XX
 PF 17-JAN-2001; 2001WO-US001340.
 XX
 PR 31-JAN-2000; 2000US-0179065P.
 PR 04-FEB-2000; 2000US-0180628P.
 PR 24-FEB-2000; 2000US-0184564P.
 PR 02-MAR-2000; 2000US-0186350P.
 PR 16-MAR-2000; 2000US-0189874P.
 PR 17-MAR-2000; 2000US-0190076P.
 PR 18-APR-2000; 2000US-0198123P.
 PR 09-MAY-2000; 2000US-0205515P.
 PR 07-JUN-2000; 2000US-0209467P.
 PR 28-JUN-2000; 2000US-0214866P.
 PR 30-JUN-2000; 2000US-0215135P.
 PR 07-JUL-2000; 2000US-0216647P.
 PR 11-JUL-2000; 2000US-0216880P.
 PR 07-JUL-2000; 2000US-0217487P.
 PR 14-JUL-2000; 2000US-0217496P.
 PR 14-JUL-2000; 2000US-0218290P.
 PR 26-JUL-2000; 2000US-0220963P.
 PR 26-JUL-2000; 2000US-0220964P.
 PR 14-AUG-2000; 2000US-0224518P.
 PR 14-AUG-2000; 2000US-0224519P.
 PR 14-AUG-2000; 2000US-0225213P.
 PR 14-AUG-2000; 2000US-0225214P.
 PR 14-AUG-2000; 2000US-0225266P.
 PR 14-AUG-2000; 2000US-0225267P.
 PR 14-AUG-2000; 2000US-0225268P.
 PR 14-AUG-2000; 2000US-0225270P.
 PR 14-AUG-2000; 2000US-0225447P.
 PR 14-AUG-2000; 2000US-0225757P.
 PR 14-AUG-2000; 2000US-0225758P.
 PR 18-AUG-2000; 2000US-0225759P.
 PR 18-AUG-2000; 2000US-0226279P.
 PR 22-AUG-2000; 2000US-0226681P.
 PR 22-AUG-2000; 2000US-0226688P.
 PR 22-AUG-2000; 2000US-0227182P.
 PR 23-AUG-2000; 2000US-0227009P.
 PR 30-AUG-2000; 2000US-0228924P.
 PR 01-SEP-2000; 2000US-0229287P.
 PR 01-SEP-2000; 2000US-0229343P.
 PR 01-SEP-2000; 2000US-0229344P.
 PR 01-SEP-2000; 2000US-0229345P.
 PR 05-SEP-2000; 2000US-0229509P.
 PR 05-SEP-2000; 2000US-0229513P.
 PR 06-SEP-2000; 2000US-0230437P.
 PR 06-SEP-2000; 2000US-0230438P.
 PR

PR 05-DEC-2000; 2000US-0251988P.
 PR 05-DEC-2000; 2000US-0256719P.
 PR 06-DEC-2000; 2000US-0251479P.
 PR 08-DEC-2000; 2000US-0251856P.
 PR 08-DEC-2000; 2000US-0251868P.
 PR 08-DEC-2000; 2000US-0251869P.
 PR 08-DEC-2000; 2000US-0251989P.
 PR 08-DEC-2000; 2000US-0251990P.
 PR 11-DEC-2000; 2000US-0254097P.
 PR 05-JAN-2001; 2001US-0259678P.
 PR 17-JAN-2001; 2001US-00764869.
 XX (HUMA-) HUMAN GENOME SCI INC.
 PA Rosen CA, Ruben SM, Barash SC;
 PI WPI; 2003-743766/70.
 XX N-PSDB; ADE45661.
 DR New cardiovascular system related polynucleotides and polypeptides,
 PT useful for preventing, treating, or ameliorating a medical condition,
 PT such as cancer of cardiovascular tissues and cancer metastases.
 XX Claim 11; SEQ ID NO 1082; 262pp; English.
 PS The invention relates to human cardiovascular system related polypeptides
 CC and the polynucleotides encoding them. The polypeptides, polynucleotides
 CC and antibodies to the polypeptides are useful for diagnosing a
 CC pathological condition or a susceptibility to a pathological condition,
 CC for preventing, treating, or ameliorating a medical condition, such as
 CC cancer of cardiovascular system tissues, proliferative disorders, foetal
 CC cancer of cardiovascular system tissues, haematopoietic disorders, diseases of
 CC the immune system, AIDS, autoimmune diseases (e.g., rheumatoid
 CC arthritis), inflammation, allergies, neurological disorders (e.g.,
 CC Alzheimer's disease, Parkinson's disease), cognitive disorders,
 CC schizophrenia, asthma, skin disorders (e.g., psoriasis), sepsis,
 CC diabetes, atherosclerosis, cardiovascular disorders, angiogenic
 CC disorders, kidney disorders, gastrointestinal disorders, pregnancy-
 CC related disorders, endocrine disorders and infections. The nucleic acids
 CC are also useful for chromosome identification, radiation hybrid mapping
 CC or long-range restriction mapping. The polypeptides and polynucleotides
 CC may also be used as food additives or preservatives to increase or
 CC decrease storage capabilities, fat content or other nutritional
 CC components. This sequence represents a human cardiovascular system
 CC related polypeptide of the invention.
 XX Sequence 99 AA;
 SQ
 Query Match 1.7%; Score 9; DB 7; Length 99;
 Best Local Similarity 100.0%; Pred. No. 3.7;
 Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 115 GEKEPSKGD 123
 DB 49 GEKEPSKGD 57
 RESULT 2S
 ABB59454
 ID ABB59454 standard; protein; 554 AA.
 XX ABB59454;
 AC ABB59454;
 XX 26-MAR-2002 (first entry)
 DT Drosophila melanogaster polypeptide SEQ ID NO 5154.
 DE Drosophila; developmental biology; cell signalling; insecticide;
 XX pharmaceutical.
 OS Drosophila melanogaster.
 XX WO200171042-A2.
 PN

XX 27-SEP-2001.
 PD 23-MAR-2001; 2001WO-US009231.
 XX 23-MAR-2000; 2000US-0191637P.
 PR 11-JUL-2000; 2000US-00614150.
 XX (PEXE) PE CORP NY.
 PA Venter JC, Adams M, Li PWD, Myers EW;
 PI WPI; 2001-656860/75.
 DR N-PSDB; ABL03557.
 XX New isolated nucleic acid detection reagent for detecting 1000 or more
 PT genes from Drosophila and for elucidating cell signaling and cell-cell
 PT interactions.
 XX Disclosure; SEQ ID NO 5154; 21pp + Sequence Listing; English.
 PS The invention relates to an isolated nucleic acid detection reagent
 CC capable of detecting 1000 or more genes from Drosophila. The invention is
 CC useful in developmental biology and in elucidating cell signalling and
 CC cell-cell interactions in higher eukaryotes for the development of
 CC insecticides, therapeutics and pharmaceutical drugs. The invention
 CC discloses genomic DNA sequences (ABL15176-ABL30511), expressed DNA
 CC sequences (ABL01840-ABL16175) and the encoded proteins (ABB57737-
 CC ABB72072). The sequence data for this patent did not form part of the
 CC printed specification, but was obtained in electronic format directly
 CC from WIPO at ftp.wipo.int/pub/published_pct_sequences
 XX Sequence 554 AA;
 SQ
 Query Match 1.7%; Score 9; DB 4; Length 554;
 Best Local Similarity 100.0%; Pred. No. 15;
 Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 248 REEEKRKE 256
 DB 488 REEEKRKE 496
 RESULT 26
 ABU44191
 ID ABU44191 standard; protein; 637 AA.
 XX ABU44191;
 AC ABU44191;
 XX 19-JUN-2003 (first entry)
 DT Protein encoded by Prokaryotic essential gene #29718.
 DE Antisense; prokaryotic essential gene; cell proliferation; drug design.
 XX Streptococcus mutans.
 OS WO200277183-A2.
 PN 03-OCT-2002.
 XX 21-MAR-2002; 2002WO-US009107.
 PF 21-MAR-2001; 2001US-00815242.
 PR 06-SEP-2001; 2001US-00948993.
 PR 25-OCT-2001; 2001US-0342923P.
 PR 08-FEB-2002; 2002US-00072851.
 PR 06-MAR-2002; 2002US-0362699P.
 XX (ELIT-) ELITRA PHARM INC.
 PA Wang L, Zamudio C, Malone C, Haselbeck R, Ohlsen KL, Zyskind JW;
 PI Wall D, Trawick JD, Carr GJ, Yamamoto R, Forsyth RA, Xu HH;

```

XX DR WPI; 2003-029926/02.
XX DR N-PSDB; ACA48061.
XX PS
XX PT New antisense nucleic acids, useful for identifying proteins or screening
XX PT for homologous nucleic acids required for cellular proliferation to
XX PT isolate candidate molecules for rational drug discovery programs.
XX PS
XX PS Claim 25; SEQ ID NO 72115; 1766bp; English.
XX CC The invention relates to an isolated nucleic acid comprising any one of
XX CC the 6213 antisense sequences given in the specification where expression
XX CC of the nucleic acid inhibits proliferation of a cell. Also included are:
XX CC (1) a vector comprising a promoter operably linked to the nucleic acid
XX CC encoding a polypeptide whose expression is inhibited by the antisense
XX CC nucleic acid; (2) a host cell containing the vector; (3) an isolated
XX CC polypeptide or its fragment whose expression is inhibited by the
XX CC antisense nucleic acid; (4) an antibody capable of specifically binding
XX CC the polypeptide; (5) producing the polypeptide; (6) inhibiting cellular
XX CC proliferation or the activity of a gene in an operon required for
XX CC proliferation; (7) identifying a compound that influences the activity of
XX CC the gene product or that has an activity against a biological pathway
XX CC required for proliferation, or that inhibits cellular proliferation; (8)
XX CC identifying a gene required for cellular proliferation or the biological
XX CC pathway in which a proliferation-required gene or its gene product lies
XX CC or a gene on which the test compound that inhibits proliferation of an
XX CC organism acts; (9) manufacturing an antibiotic; (10) profiling a
XX CC compound's activity; (11) a culture comprising strains in which the gene
XX CC product is overexpressed or underexpressed; (12) determining the extent
XX CC to which each of the strains is present in a culture or collection of
XX CC strains; or (13) identifying the target of a compound that inhibits the
XX CC proliferation of an organism. The antisense nucleic acids are useful for
XX CC identifying proteins or screening for homologous nucleic acids required
XX CC for cellular proliferation to isolate candidate molecules for rational
XX CC drug discovery programs, or for screening homologous nucleic acids
XX CC required for proliferation in cells other than S. aureus, S. typhimurium,
XX CC K. pneumoniae or P. aeruginosa. The present sequence is encoded by one of
XX CC the target prokaryotic essential genes. Note: The sequence data for this
XX CC patent did not form part of the printed specification, but was obtained
XX CC in electronic format directly from WIPO at
XX CC ftp.wipo.int/pub/published_pct_sequences
XX PS
XX PS Sequence 637 AA;
XX PS
XX PS Query Match 1.7%; Score 9; DB 6; Length 637;
XX PS Best Local Similarity 100.0%; Pred. No. 17;
XX PS Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
XX PS
XX PS 57 BELSRQLED 65
XX PS |||||
XX PS 629 BELSRQLED 637
XX PS
XX PS RESULT 27
XX PS AAR89275
XX PS ID AAR89275 standard; protein; 729 AA.
XX PS AC
XX PS AAR89275;
XX PS
XX PS 25-JUL-1996 (first entry)
XX PS
XX PS Yeast coagulation protein FLO8.
XX PS
XX PS Saccharomyces cerevisiae; coagulation; FLO8; mutant; wild type; wine;
XX PS beer; yeast; flavour; alcoholic beverage.
XX PS
XX PS Saccharomyces cerevisiae.
XX PS
XX PS JP08000270-A.
XX PS
XX PS 09-JAN-1996.
XX PS
XX PS 23-JUN-1994; 94JP-00141520.
XX PS
XX PS 23-JUN-1994; 94JP-00141520.
XX PS (KIRI ) KIRIN BREWERY KK.
XX PS
XX PS WPI; 1996-091654/10.
XX PS N-PSDB; AAR9257.
XX PS
XX PS Yeast coagulating gene FLO8 and prepn. of yeast with increased
XX PS coagulation - also prepn. of yeast with decreased coagulation, both
XX PS useful in wine making and brewing industries.
XX PS
XX PS Claim 1; Page 13-16; 18pp; Japanese.
XX PS
XX PS This is the amino acid sequence of the Saccharomyces cerevisiae
XX PS coagulation protein FLO8. The corresp. gene was isolated by screening a
XX PS non-coagulating yeast strain (YPH500) transformed with a library
XX PS generated from DNA from a coagulating yeast strain (ARCC-60715). The full
XX PS length gene can be used to generate mutants having reduced or increased
XX PS coagulation activity. The strain carrying the wild type or mutant FLO8
XX PS genes can be used in wine and beer making, where the coagulating activity
XX PS of the yeast can affect the flavour of the alcoholic beverage
XX PS
XX PS Sequence 729 AA;
XX PS
XX PS Query Match 1.7%; Score 9; DB 2; Length 729;
XX PS Best Local Similarity 100.0%; Pred. No. 19;
XX PS Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
XX PS
XX PS 164 NTLSTPEEK 172
XX PS |||||
XX PS 601 NTLSTPEEK 609
XX PS
XX PS RESULT 28
XX PS AAY12665
XX PS ID AAY12665 standard; protein; 65 AA.
XX PS AC
XX PS AAY12665;
XX PS
XX PS 22-JUN-1999 (first entry)
XX PS
XX PS Human 5' EST secreted protein SEQ ID NO: 330 from WO 9906553.
XX PS
XX PS Human; secreted protein; EST; expressed sequence tag; diagnosis;
XX PS forensic; gene therapy; chromosome mapping; signal peptide;
XX PS upstream regulatory sequence; cytokine activity; cell proliferation;
XX PS differentiation; haematopoiesis regulation; tissue growth regulation;
XX PS reproductive hormone regulation; chemotactic; chemokinetic; haemostatic;
XX PS thrombolytic; antiinflammatory; tumour inhibition; antitumour.
XX PS
XX PS Homo sapiens.
XX PS
XX PS WO9906553-A2.
XX PS
XX PS 11-FEB-1999.
XX PS
XX PS 31-JUL-1998; 98WO-IB001237.
XX PS
XX PS 01-AUG-1997; 97US-00905051.
XX PS
XX PS (CBST ) GENSET.
XX PS
XX PS Dumas Milne Edwards J, Duclert A, Lacroix B;
XX PS
XX PS WPI; 1999-153783/13.
XX PS N-PSDB; AAX41523.
XX PS
XX PS New nucleic acids encoding human secreted proteins - obtained from cDNA
XX PS libraries derived from umbilical cord, lymph ganglia, lymphocytes and
XX PS placental tissue.
XX PS
XX PS Claim 34; Page 408; 411pp; English.

```

XX AAX41379 to AAX41526 represent 5' expressed sequence tags (ESTs) for
CC human secreted proteins, and encode the proteins given in AAY12521 to
CC AAY12668, respectively. The proteins given represent the signal peptide
CC and an N-terminal fragment of a secreted protein. The nucleic acid
CC sequences can be used for producing secreted human gene products. They
CC can also be used to develop products for diagnosis and therapy. The
CC proteins obtained may have cytokine activity, cell
CC proliferation/differentiation activity, haematopoiesis regulating
CC activity, tissue growth regulating activity, reproductive hormone
CC regulating activity, chemotactic/chemokinetic activity, haemostatic and
CC thrombolytic activity, receptor/ligand activity, antiinflammatory
CC activity, tumour inhibition activity or other activities. The products
CC can be used in forensic, gene therapy and chromosome mapping procedures.
CC The sequences can also be used for obtaining corresponding promoter
CC sequences. The nucleic acids encoding the signal peptide can be used for
CC directing extracellular secretion of a polypeptide or the insertion of a
CC polypeptide into a membrane, or importing a polypeptide into a cell
XX
SQ Sequence 65 AA;
Query Match 1.5%; Score 8; DB 2; Length 65;
Best Local Similarity 100.0%; Pred. No. 24;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 492 PGQAQPS 499
Db 10 PGQAQPS 17
|||||
RESULT 29
AAY01195
ID AAY01195 standard; protein; 156 AA.
XX
AC AAY01195;
XX
XX 18-MAY-1999 (first entry)
DT Polypeptide fragment encoded by gene 15.
DE
DE Human; secreted protein; gene therapy; protein therapy; tissue; cancer;
XX tumour; neurodegenerative disorder; leukaemia; autoimmune disease; AIDS;
XX developmental abnormality; foetal deficiency; Alzheimer's disease;
XX cognitive disorder; schizophrenia; immunological disorder; mood disorder;
XX immune deficiency disease; respiratory disorder; arthritis; skeletal;
XX haematopoietic disorder; neural; osteoporosis; metabolic disorders;
XX cardiovascular; endocrine; gastrointestinal; asthma; diagnosis.
XX Homo sapiens.
CS
OS WO9901020-A2.
XX
XX 14-JAN-1999.
XX
XX 30-JUN-1998; 98WO-US013608.
XX
XX 01-JUL-1997; 97US-0051381P.
XX 01-JUL-1997; 97US-0051480P.
XX 12-SEP-1997; 97US-0058598P.
XX 12-SEP-1997; 97US-0058663P.
XX
XX (HUMA-) HUMAN GENOME SCI INC.
XX
XX Carter KC, Feng P, Rosen CA, Ruben SM, Endress GA;
XX N-PSDB; AAX22125.
XX
XX WPI; 1999-105683/09.
XX
XX
XX
XX New isolated human genes and the secreted polypeptides they encode -
XX useful for diagnosis and treatment of e.g. cancers, neurological
XX disorders, immune diseases, immune deficiency diseases or blood
XX disorders.
XX

PS Disclosure; Page 26; 179pp; English.
XX
CC The invention relates to nucleic acid sequences (AAX22111 to AAX22134)
CC encoding human secreted proteins (AAY01135 to AAY01158). The secreted
CC protein gene sequences are deposited with the ATCC under deposit number
CC ATCC 29918. Host cells comprising recombinant vectors containing the
CC nucleic acid sequences are used for the recombinant production of the
CC secreted proteins. The polynucleotide and amino acid sequences are useful
CC for are useful for preventing, treating or ameliorating medical
CC conditions e.g. by protein or gene therapy. Pathological conditions can
CC be also diagnosed by determining the amount of the new polypeptides in a
CC sample or by determining the presence of mutations in the new
CC polynucleotides. Specific uses are described for each of the
CC polynucleotides, based on which tissues they are most highly expressed
CC in, and include developing products for the diagnosis or treatment of
CC cancer, tumours, developmental abnormalities and foetal deficiencies,
CC autoimmune diseases, lymphomas, Alzheimer's and cognitive disorders,
CC schizophrenia, immunological disorders, immune deficiency diseases
CC (AIDS), mood disorders, respiratory disorders, arthritis, asthma,
CC haematopoietic disorders, neural disorders, skeletal disorders,
CC osteoporosis, metabolic disorders, cardiovascular disorders, endocrine
CC disorders or gastrointestinal disorders. The polypeptides are also useful
CC for identifying their binding partners. The present sequence represents a
CC polypeptide fragment encoded by a gene of the invention (see descriptor
CC line for gene number)
XX
SQ Sequence 156 AA;
Query Match 1.5%; Score 8; DB 2; Length 156;
Best Local Similarity 100.0%; Pred. No. 50;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 468 RVQDLSAG 475
Db 87 RVQDLSAG 94
|||||
RESULT 30
AAG08727
ID AAG08727 standard; protein; 170 AA.
XX
AC AAG08727;
XX
XX 17-OCT-2000 (first entry)
DT Arabidopsis thaliana protein fragment SEQ ID NO: 6379.
DE
DE Protein identification; signal transduction pathway; metabolic pathway;
XX hybridisation assay; genetic mapping; gene expression control; promoter;
XX termination sequence.
XX Arabidopsis thaliana.
XX BP1033405-A2.
XX
XX 06-SEP-2000.
XX
XX 25-FEB-2000; 2000EP-00301439.
XX
XX 25-FEB-1999; 99US-0121825P.
XX 05-MAR-1999; 99US-0123180P.
XX 09-MAR-1999; 99US-0123548P.
XX 23-MAR-1999; 99US-0125789P.
XX 25-MAR-1999; 99US-0126264P.
XX 29-MAR-1999; 99US-0126783P.
XX 01-APR-1999; 99US-0127462P.
XX 06-APR-1999; 99US-0128234P.
XX 08-APR-1999; 99US-0128714P.
XX 16-APR-1999; 99US-0129845P.
XX 19-APR-1999; 99US-0130077P.
XX 21-APR-1999; 99US-0130443P.
XX 23-APR-1999; 99US-0130510P.
XX 23-APR-1999; 99US-0130891P.

PR 28-APR-1999; 99US-0131449P.
PR 30-APR-1999; 99US-0132048P.
PR 30-APR-1999; 99US-0132407P.
PR 04-MAY-1999; 99US-0132484P.
PR 05-MAY-1999; 99US-0132485P.
PR 06-MAY-1999; 99US-0132486P.
PR 06-MAY-1999; 99US-0132487P.
PR 07-MAY-1999; 99US-0132863P.
PR 11-MAY-1999; 99US-0132866P.
PR 14-MAY-1999; 99US-0134218P.
PR 14-MAY-1999; 99US-0134219P.
PR 14-MAY-1999; 99US-0134221P.
PR 14-MAY-1999; 99US-0134370P.
PR 18-MAY-1999; 99US-0134768P.
PR 19-MAY-1999; 99US-0134941P.
PR 20-MAY-1999; 99US-0135124P.
PR 21-MAY-1999; 99US-0135353P.
PR 24-MAY-1999; 99US-0135629P.
PR 25-MAY-1999; 99US-0136021P.
PR 27-MAY-1999; 99US-0136392P.
PR 28-MAY-1999; 99US-0136782P.
PR 01-JUN-1999; 99US-0137222P.
PR 03-JUN-1999; 99US-0137528P.
PR 04-JUN-1999; 99US-0137502P.
PR 07-JUN-1999; 99US-0137724P.
PR 08-JUN-1999; 99US-0138094P.
PR 10-JUN-1999; 99US-0138540P.
PR 10-JUN-1999; 99US-0138847P.
PR 14-JUN-1999; 99US-0139113P.
PR 16-JUN-1999; 99US-0139452P.
PR 16-JUN-1999; 99US-0139453P.
PR 17-JUN-1999; 99US-0139492P.
PR 18-JUN-1999; 99US-0139454P.
PR 18-JUN-1999; 99US-0139455P.
PR 18-JUN-1999; 99US-0139456P.
PR 18-JUN-1999; 99US-0139457P.
PR 18-JUN-1999; 99US-0139458P.
PR 18-JUN-1999; 99US-0139459P.
PR 18-JUN-1999; 99US-0139460P.
PR 18-JUN-1999; 99US-0139461P.
PR 18-JUN-1999; 99US-0139462P.
PR 18-JUN-1999; 99US-0139463P.
PR 18-JUN-1999; 99US-0139750P.
PR 18-JUN-1999; 99US-0139763P.
PR 21-JUN-1999; 99US-0139817P.
PR 22-JUN-1999; 99US-0139899P.
PR 23-JUN-1999; 99US-0140353P.
PR 23-JUN-1999; 99US-0140354P.
PR 24-JUN-1999; 99US-0140698P.
PR 28-JUN-1999; 99US-0140823P.
PR 29-JUN-1999; 99US-0140991P.
PR 30-JUN-1999; 99US-0141287P.
PR 01-JUL-1999; 99US-0141842P.
PR 01-JUL-1999; 99US-0142154P.
PR 02-JUL-1999; 99US-0142055P.
PR 06-JUL-1999; 99US-0142390P.
PR 08-JUL-1999; 99US-0142803P.
PR 09-JUL-1999; 99US-0142920P.
PR 12-JUL-1999; 99US-0142977P.
PR 13-JUL-1999; 99US-0143542P.
PR 14-JUL-1999; 99US-0143624P.
PR 15-JUL-1999; 99US-0144005P.
PR 16-JUL-1999; 99US-0144085P.
PR 16-JUL-1999; 99US-0144086P.
PR 19-JUL-1999; 99US-0144325P.
PR 19-JUL-1999; 99US-0144331P.
PR 19-JUL-1999; 99US-0144332P.
PR 19-JUL-1999; 99US-0144333P.
PR 19-JUL-1999; 99US-0144334P.
PR 19-JUL-1999; 99US-0144335P.
PR 20-JUL-1999; 99US-0144352P.
PR 20-JUL-1999; 99US-0144632P.
PR 20-JUL-1999; 99US-0144684P.
PR 20-JUL-1999; 99US-0144884P.
PR 21-JUL-1999; 99US-0144814P.
PR 21-JUL-1999; 99US-0145086P.
PR 22-JUL-1999; 99US-0145088P.
PR 22-JUL-1999; 99US-0145085P.
PR 22-JUL-1999; 99US-0145087P.
PR 22-JUL-1999; 99US-0145089P.
PR 22-JUL-1999; 99US-0145192P.
PR 23-JUL-1999; 99US-0145145P.
PR 23-JUL-1999; 99US-0145218P.
PR 23-JUL-1999; 99US-0145224P.
PR 26-JUL-1999; 99US-0145276P.
PR 27-JUL-1999; 99US-0145913P.
PR 27-JUL-1999; 99US-0145918P.
PR 28-JUL-1999; 99US-0145919P.
PR 28-JUL-1999; 99US-0145951P.
PR 02-AUG-1999; 99US-0146386P.
PR 02-AUG-1999; 99US-0146388P.
PR 02-AUG-1999; 99US-0146389P.
PR 03-AUG-1999; 99US-0147038P.
PR 04-AUG-1999; 99US-0147204P.
PR 04-AUG-1999; 99US-0147302P.
PR 05-AUG-1999; 99US-0147192P.
PR 05-AUG-1999; 99US-0147260P.
PR 06-AUG-1999; 99US-0147303P.
PR 06-AUG-1999; 99US-0147416P.
PR 09-AUG-1999; 99US-0147493P.
PR 09-AUG-1999; 99US-0147935P.
PR 10-AUG-1999; 99US-0148171P.
PR 11-AUG-1999; 99US-0148319P.
PR 12-AUG-1999; 99US-0148341P.
PR 13-AUG-1999; 99US-0148565P.
PR 13-AUG-1999; 99US-0148684P.
PR 16-AUG-1999; 99US-0149368P.
PR 17-AUG-1999; 99US-0149175P.
PR 18-AUG-1999; 99US-0149426P.
PR 20-AUG-1999; 99US-0149722P.
PR 20-AUG-1999; 99US-0149723P.
PR 20-AUG-1999; 99US-0149829P.
PR 23-AUG-1999; 99US-0149902P.
PR 23-AUG-1999; 99US-0149930P.
PR 25-AUG-1999; 99US-0150566P.
PR 26-AUG-1999; 99US-0150884P.
PR 26-AUG-1999; 99US-0151065P.
PR 27-AUG-1999; 99US-0151066P.
PR 27-AUG-1999; 99US-0151080P.
PR 30-AUG-1999; 99US-0151303P.
PR 31-AUG-1999; 99US-0151438P.
PR 01-SEP-1999; 99US-0151930P.
PR 07-SEP-1999; 99US-0152363P.
PR 10-SEP-1999; 99US-0153070P.
PR 13-SEP-1999; 99US-0153758P.
PR 15-SEP-1999; 99US-0154018P.
PR 16-SEP-1999; 99US-0154039P.
PR 20-SEP-1999; 99US-0154779P.
PR 22-SEP-1999; 99US-0155139P.
PR 23-SEP-1999; 99US-0155486P.
PR 24-SEP-1999; 99US-0155659P.
PR 28-SEP-1999; 99US-0156458P.
PR 29-SEP-1999; 99US-0156596P.
PR 04-OCT-1999; 99US-0157117P.
PR 05-OCT-1999; 99US-0157753P.
PR 06-OCT-1999; 99US-0157865P.
PR 07-OCT-1999; 99US-0158029P.
PR 08-OCT-1999; 99US-0158232P.
PR 12-OCT-1999; 99US-0158369P.
PR 13-OCT-1999; 99US-0159293P.
PR 13-OCT-1999; 99US-0159295P.
PR 14-OCT-1999; 99US-0159329P.
PR 14-OCT-1999; 99US-0159330P.
PR 14-OCT-1999; 99US-0159331P.
PR 14-OCT-1999; 99US-0159637P.
PR 14-OCT-1999; 99US-0159638P.

PR 23-JUL-1999; 99US-0145145P.
 PR 23-JUL-1999; 99US-0145218P.
 PR 23-JUL-1999; 99US-0145222P.
 PR 26-JUL-1999; 99US-0145276P.
 PR 27-JUL-1999; 99US-0145913P.
 PR 27-JUL-1999; 99US-0145918P.
 PR 27-JUL-1999; 99US-0145919P.
 PR 28-JUL-1999; 99US-0145951P.
 PR 02-AUG-1999; 99US-0146386P.
 PR 02-AUG-1999; 99US-0146388P.
 PR 02-AUG-1999; 99US-0146389P.
 PR 03-AUG-1999; 99US-0147038P.
 PR 04-AUG-1999; 99US-0147204P.
 PR 05-AUG-1999; 99US-0147302P.
 PR 05-AUG-1999; 99US-0147192P.
 PR 05-AUG-1999; 99US-0147260P.
 PR 06-AUG-1999; 99US-0147303P.
 PR 06-AUG-1999; 99US-0147416P.
 PR 09-AUG-1999; 99US-0147493P.
 PR 09-AUG-1999; 99US-0147935P.
 PR 10-AUG-1999; 99US-0148171P.
 PR 11-AUG-1999; 99US-0148319P.
 PR 12-AUG-1999; 99US-0148341P.
 PR 13-AUG-1999; 99US-0148565P.
 PR 13-AUG-1999; 99US-0148684P.
 PR 16-AUG-1999; 99US-0149368P.
 PR 17-AUG-1999; 99US-0149175P.
 PR 18-AUG-1999; 99US-0149426P.
 PR 20-AUG-1999; 99US-0149722P.
 PR 20-AUG-1999; 99US-0149723P.
 PR 20-AUG-1999; 99US-0149929P.
 PR 23-AUG-1999; 99US-0149902P.
 PR 23-AUG-1999; 99US-0149930P.
 PR 25-AUG-1999; 99US-0150566P.
 PR 26-AUG-1999; 99US-0150884P.
 PR 27-AUG-1999; 99US-0151065P.
 PR 27-AUG-1999; 99US-0151066P.
 PR 27-AUG-1999; 99US-0151080P.
 PR 30-AUG-1999; 99US-0151303P.
 PR 31-AUG-1999; 99US-0151438P.
 PR 01-SEP-1999; 99US-0151930P.
 PR 07-SEP-1999; 99US-0152363P.
 PR 10-SEP-1999; 99US-0153070P.
 PR 13-SEP-1999; 99US-0153758P.
 PR 15-SEP-1999; 99US-0154018P.
 PR 16-SEP-1999; 99US-0154039P.
 PR 20-SEP-1999; 99US-0154779P.
 PR 22-SEP-1999; 99US-0155139P.
 PR 23-SEP-1999; 99US-0155486P.
 PR 24-SEP-1999; 99US-0156599P.
 PR 28-SEP-1999; 99US-0156458P.
 PR 29-SEP-1999; 99US-015659P.
 PR 04-OCT-1999; 99US-0157117P.
 PR 05-OCT-1999; 99US-015753P.
 PR 06-OCT-1999; 99US-0157865P.
 PR 07-OCT-1999; 99US-0158029P.
 PR 08-OCT-1999; 99US-0158232P.
 PR 12-OCT-1999; 99US-0158369P.
 PR 13-OCT-1999; 99US-0159293P.
 PR 13-OCT-1999; 99US-0159294P.
 PR 13-OCT-1999; 99US-0159295P.
 PR 14-OCT-1999; 99US-0159329P.
 PR 14-OCT-1999; 99US-0159330P.
 PR 14-OCT-1999; 99US-0159331P.
 PR 14-OCT-1999; 99US-0159637P.
 PR 14-OCT-1999; 99US-0159638P.
 PR 18-OCT-1999; 99US-0159584P.
 PR 21-OCT-1999; 99US-0160741P.
 PR 21-OCT-1999; 99US-0160767P.
 PR 21-OCT-1999; 99US-0160768P.
 PR 21-OCT-1999; 99US-0160770P.
 PR 21-OCT-1999; 99US-0160814P.
 PR 21-OCT-1999; 99US-0160815P.

PR 22-OCT-1999; 99US-0160980P.
 PR 22-OCT-1999; 99US-0160981P.
 PR 22-OCT-1999; 99US-0160989P.
 PR 23-OCT-1999; 99US-0161404P.
 PR 23-OCT-1999; 99US-0161405P.
 PR 23-OCT-1999; 99US-0161406P.
 PR 23-OCT-1999; 99US-0161359P.
 PR 26-OCT-1999; 99US-0161360P.
 PR 26-OCT-1999; 99US-0161361P.
 PR 26-OCT-1999; 99US-0161920P.
 PR 28-OCT-1999; 99US-0161922P.
 PR 28-OCT-1999; 99US-0161933P.
 PR 29-OCT-1999; 99US-0162142P.

Query Match 1.5%; Score 8; DB 3; Length 174;
 Best Local Similarity 100.0%; Pred. No. 54;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 21 AAPAAVEAE 28
 |||||
 DB 34 AAPAAVEAE 41

RESULT 32

ABG19843
 ID ABG19843 standard; protein; 194 AA.

XX AC ABG19843;

XX DT 18-FEB-2002 (first entry)

XX DE Novel human diagnostic protein #19834.

XX Human; chromosome mapping; gene mapping; gene therapy; forensic;
 XX food supplement; medical imaging; diagnostic; genetic disorder.

XX OS Homo sapiens.

XX PN WC200175067-A2.

XX PD 11-OCT-2001.

XX PF 30-MAR-2001; 2001WO-US008631.

XX PR 31-MAR-2000; 2000US-00540217.

XX PR 23-AUG-2000; 2000US-00649167.

XX PA (HYSE-) HYSEQ INC.

XX PI Drmanac RT, Liu C, Tang YT;

XX DR WPI: 2001-639362/73.

XX DR N-PSDB; AAS84030.

XX PT New isolated polynucleotide and encoded polypeptides, useful in
 PT diagnostics, forensics, gene mapping, identification of mutations
 PT responsible for genetic disorders or other traits and to assess
 PT biodiversity.

XX PS Claim 20; SEQ ID NO 50202; 103pp; English.

XX CC The invention relates to isolated polynucleotide (I) and polypeptide (II)
 CC sequences. (I) is useful as hybridisation probes, polymerase chain
 CC reaction (PCR) primers, oligomers, and for chromosome and gene mapping,
 CC and in recombinant production of (II). The polynucleotides are also used
 CC in diagnostics as expressed sequence tags for identifying expressed
 CC genes. (I) is useful in gene therapy techniques to restore normal
 CC activity of (II) or to treat disease states involving (II). (II) is
 CC useful for generating antibodies against it, detecting or quantitating a
 CC polypeptide in tissue, as molecular weight markers and as a food
 CC supplement. (II) and its binding partners are useful in medical imaging
 CC of sites expressing (II). (I) and (II) are useful for treating disorders
 CC involving aberrant protein expression or biological activity. The

CC polypeptide and polynucleotide sequences have applications in
CC diagnostics, forensics, gene mapping, identification of mutations
CC responsible for genetic disorders or other traits to assess biodiversity
CC and to produce other types of data and products dependent on DNA and
CC amino acid sequences. ABG0010-ABG30377 represent novel human diagnostic
CC amino acid sequences of the invention. Note: The sequence data for this
CC patent did not appear in the printed specification, but was obtained in
CC electronic format directly from WIPO at
CC ftp.wipo.int/pub/published_pct_sequences
XX
SQ Sequence 194 AA;

Query Match 1.5%; Score 8; DB 4; Length 194;
Best Local Similarity 100.0%; Pred. No. 60;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 476 GQGLTDS 483
Db 136 GQGLTDS 143
|||||

RESULT 33
AAG20570
ID AAG20570 standard; protein; 224 AA.

XX AC AAG20570;

XX DT 17-OCT-2000 (first entry)

XX DE Arabidopsis thaliana protein fragment SEQ ID NO: 22814.

XX KW Protein identification; signal transduction pathway; metabolic pathway;
XX hybridisation assay; genetic mapping; gene expression control; promoter;
XX termination sequence.

XX OS Arabidopsis thaliana.

XX PN EPI033405-A2.

XX PD 06-SEP-2000.

XX PF 25-FEB-2000; 2000EP-0301439.

XX PR 25-FEB-1999; 99US-0121825P.

XX PR 05-MAR-1999; 99US-0123180P.

XX PR 09-MAR-1999; 99US-0123548P.

XX PR 23-MAR-1999; 99US-0125788P.

XX PR 25-MAR-1999; 99US-0126284P.

XX PR 29-MAR-1999; 99US-0126785P.

XX PR 01-APR-1999; 99US-0127462P.

XX PR 06-APR-1999; 99US-0128234P.

XX PR 08-APR-1999; 99US-0128714P.

XX PR 16-APR-1999; 99US-0129845P.

XX PR 19-APR-1999; 99US-0130077P.

XX PR 21-APR-1999; 99US-0130449P.

XX PR 23-APR-1999; 99US-0130510P.

XX PR 23-APR-1999; 99US-0130891P.

XX PR 28-APR-1999; 99US-0131449P.

XX PR 30-APR-1999; 99US-0132048P.

XX PR 30-APR-1999; 99US-0132407P.

XX PR 04-MAY-1999; 99US-0132484P.

XX PR 05-MAY-1999; 99US-0132485P.

PR 21-MAY-1999; 99US-0135353P.
PR 24-MAY-1999; 99US-0135629P.
PR 25-MAY-1999; 99US-0136021P.
PR 27-MAY-1999; 99US-0136392P.
PR 28-MAY-1999; 99US-0136782P.
PR 01-JUN-1999; 99US-0137222P.
PR 03-JUN-1999; 99US-0137528P.
PR 04-JUN-1999; 99US-0137502P.
PR 07-JUN-1999; 99US-0137724P.
PR 08-JUN-1999; 99US-0138094P.
PR 10-JUN-1999; 99US-0138540P.
PR 10-JUN-1999; 99US-0138847P.
PR 14-JUN-1999; 99US-0139119P.
PR 16-JUN-1999; 99US-0139452P.
PR 16-JUN-1999; 99US-0139453P.
PR 17-JUN-1999; 99US-0139492P.
PR 18-JUN-1999; 99US-0139454P.
PR 18-JUN-1999; 99US-0139455P.
PR 18-JUN-1999; 99US-0139456P.
PR 18-JUN-1999; 99US-0139457P.
PR 18-JUN-1999; 99US-0139458P.
PR 18-JUN-1999; 99US-0139459P.
PR 18-JUN-1999; 99US-0139460P.
PR 18-JUN-1999; 99US-0139461P.
PR 18-JUN-1999; 99US-0139462P.
PR 18-JUN-1999; 99US-0139463P.
PR 18-JUN-1999; 99US-0139750P.
PR 18-JUN-1999; 99US-0139763P.
PR 21-JUN-1999; 99US-0139817P.
PR 22-JUN-1999; 99US-0139899P.
PR 23-JUN-1999; 99US-0140353P.
PR 23-JUN-1999; 99US-0140354P.
PR 24-JUN-1999; 99US-0140695P.
PR 28-JUN-1999; 99US-0140823P.
PR 29-JUN-1999; 99US-0140991P.
PR 30-JUN-1999; 99US-0141287P.
PR 01-JUL-1999; 99US-0141842P.
PR 01-JUL-1999; 99US-0142154P.
PR 02-JUL-1999; 99US-0142055P.
PR 06-JUL-1999; 99US-0142390P.
PR 08-JUL-1999; 99US-0142803P.
PR 09-JUL-1999; 99US-0142920P.
PR 12-JUL-1999; 99US-0142977P.
PR 13-JUL-1999; 99US-0143542P.
PR 14-JUL-1999; 99US-0143624P.
PR 15-JUL-1999; 99US-0144005P.
PR 16-JUL-1999; 99US-0144085P.
PR 16-JUL-1999; 99US-0144086P.
PR 19-JUL-1999; 99US-0144325P.
PR 19-JUL-1999; 99US-0144331P.
PR 19-JUL-1999; 99US-0144332P.
PR 19-JUL-1999; 99US-0144333P.
PR 19-JUL-1999; 99US-0144334P.
PR 19-JUL-1999; 99US-0144335P.
PR 20-JUL-1999; 99US-0144352P.
PR 20-JUL-1999; 99US-0144632P.
PR 20-JUL-1999; 99US-0144884P.
PR 21-JUL-1999; 99US-0144814P.
PR 21-JUL-1999; 99US-0145086P.
PR 21-JUL-1999; 99US-0145088P.
PR 22-JUL-1999; 99US-0145085P.
PR 22-JUL-1999; 99US-0145087P.
PR 22-JUL-1999; 99US-0145089P.
PR 22-JUL-1999; 99US-0145192P.
PR 23-JUL-1999; 99US-0145145P.
PR 23-JUL-1999; 99US-0145218P.
PR 23-JUL-1999; 99US-0145224P.
PR 26-JUL-1999; 99US-0145276P.
PR 27-JUL-1999; 99US-0145913P.
PR 27-JUL-1999; 99US-0145918P.
PR 27-JUL-1999; 99US-0145919P.
PR 28-JUL-1999; 99US-0145951P.
PR 02-AUG-1999; 99US-0146386P.

```
PR 02-AUG-1999; 99US-0146388P.
PR 02-AUG-1999; 99US-0146389P.
PR 03-AUG-1999; 99US-0147038P.
PR 04-AUG-1999; 99US-0147204P.
PR 04-AUG-1999; 99US-0147302P.
PR 05-AUG-1999; 99US-0147192P.
PR 05-AUG-1999; 99US-0147260P.
PR 06-AUG-1999; 99US-0147303P.
PR 06-AUG-1999; 99US-0147416P.
PR 09-AUG-1999; 99US-0147493P.
PR 09-AUG-1999; 99US-0147935P.
PR 10-AUG-1999; 99US-0148171P.
PR 11-AUG-1999; 99US-0148319P.
PR 12-AUG-1999; 99US-0148341P.
PR 13-AUG-1999; 99US-0148565P.
PR 13-AUG-1999; 99US-0148684P.
PR 16-AUG-1999; 99US-0149368P.
PR 17-AUG-1999; 99US-0149175P.
PR 18-AUG-1999; 99US-0149426P.
PR 20-AUG-1999; 99US-0149722P.
PR 20-AUG-1999; 99US-0149723P.
PR 20-AUG-1999; 99US-0149929P.
PR 23-AUG-1999; 99US-0149902P.
PR 23-AUG-1999; 99US-0149930P.
PR 25-AUG-1999; 99US-0150566P.
PR 26-AUG-1999; 99US-0150884P.
PR 27-AUG-1999; 99US-0151065P.
PR 27-AUG-1999; 99US-0151066P.
PR 27-AUG-1999; 99US-0151080P.
PR 30-AUG-1999; 99US-0151303P.
PR 31-AUG-1999; 99US-0151438P.
PR 01-SEP-1999; 99US-0151930P.
PR 01-SEP-1999; 99US-0152363P.
PR 10-SEP-1999; 99US-0153070P.
PR 13-SEP-1999; 99US-0153758P.
PR 15-SEP-1999; 99US-0154018P.
PR 16-SEP-1999; 99US-0154039P.
PR 20-SEP-1999; 99US-0154779P.
PR 22-SEP-1999; 99US-0155139P.
PR 23-SEP-1999; 99US-0155486P.
PR 24-SEP-1999; 99US-0155659P.
PR 28-SEP-1999; 99US-0156458P.
PR 29-SEP-1999; 99US-0156596P.
PR 04-OCT-1999; 99US-0157117P.
PR 05-OCT-1999; 99US-0157553P.
PR 06-OCT-1999; 99US-0157865P.
PR 07-OCT-1999; 99US-0158029P.
PR 08-OCT-1999; 99US-0158232P.
PR 12-OCT-1999; 99US-0158369P.
PR 13-OCT-1999; 99US-0159293P.
PR 13-OCT-1999; 99US-0159294P.
PR 13-OCT-1999; 99US-0159295P.
PR 14-OCT-1999; 99US-0159329P.
PR 14-OCT-1999; 99US-0159330P.
PR 14-OCT-1999; 99US-0159331P.
PR 14-OCT-1999; 99US-0159637P.
PR 14-OCT-1999; 99US-0159638P.
PR 18-OCT-1999; 99US-0159584P.
PR 21-OCT-1999; 99US-0160741P.
PR 21-OCT-1999; 99US-0160767P.
PR 21-OCT-1999; 99US-0160768P.
PR 21-OCT-1999; 99US-0160770P.
PR 21-OCT-1999; 99US-0160814P.
PR 21-OCT-1999; 99US-0160815P.
PR 22-OCT-1999; 99US-0160980P.
PR 22-OCT-1999; 99US-0160981P.
PR 22-OCT-1999; 99US-0160989P.
PR 25-OCT-1999; 99US-0161404P.
PR 25-OCT-1999; 99US-0161405P.
PR 25-OCT-1999; 99US-0161408P.
PR 26-OCT-1999; 99US-0161359P.
PR 26-OCT-1999; 99US-0161360P.
PR 26-OCT-1999; 99US-0161361P.

PR 28-OCT-1999; 99US-0161920P.
PR 28-OCT-1999; 99US-0161922P.
PR 28-OCT-1999; 99US-0161933P.
PR 29-OCT-1999; 99US-0162142P.

Query Match 1.5%; Score 8; DB 3; Length 224;
Best Local Similarity 100.0%; Pred. No. 67;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 21 AAPAVEAE 28
Db 84 AAPAVEAE 91

RESULT 34
ABU05665
ID ABU05665 standard; protein; 229 AA.
XX AC ABU05665;
XX DT 08-APR-2003 (first entry)
XX DE M. tuberculosis and M. leprae marker protein #316.
XX KW Mycobacterioses; survival; virulence; protective antigen; vaccine;
XX KW Mycobacterial disease; tuberculosis; leprosy.
XX OS Mycobacterium tuberculosis.
XX OS Mycobacterium leprae.
XX FN WO200274903-A2.
XX PD 26-SEP-2002.
XX PP 22-FEB-2002; 2002WO-IB001973.
XX PR 22-FEB-2001; 2001US-0270123P.
XX PA (INSP ) INST PASTEUR.
XX PI Cole S;
XX DR WPI; 2002-759985/82.
XX PT Identifying and selecting genes for survival or virulence of mycobacteria
PT by a comparative genomic analysis of the sequences of Mycobacterium
PT tuberculosis and M. leprae.
XX PS Claim 17; Page 500-501; 874pp; English.
XX CC This invention relates to a novel method for identifying essential genes
CC for survival or virulence of mycobacteria species. The method comprises
CC aligning the genomic sequence of a first mycobacterium species on a
CC genomic sequence of a second mycobacterium species and selecting a
CC polynucleotide sequence that is highly conserved in both genomes with no
CC counterparts in other bacterial genomic sequences and that corresponds to
CC an essential gene for the survival or virulence of mycobacterium species.
CC The method of the invention is useful for detecting M. tuberculosis or M.
CC leprae infection. The method reduces the number of potential new targets
CC and protective antigens for new drugs and vaccine compositions to treat
CC and prevent mycobacterial diseases, particularly tuberculosis and
CC leprosy. The present sequence represents a marker protein from
CC Mycobacterium tuberculosis and Mycobacterium leprae identified using the
CC method of the invention
XX SQ Sequence 229 AA;

Query Match 1.5%; Score 8; DB 5; Length 229;
Best Local Similarity 100.0%; Pred. No. 68;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 18 PSQAPAV 25
|||||
```


CC The invention relates to 971 novel human cDNA sequences (ADC29919-
 CC ADC30889) and the polypeptides they encode (ADC30890-ADC31860). The
 CC invention also relates to nucleic acid sequences over 99% identical with
 CC the novel human cDNAs. The invention additionally encompasses expression
 CC vectors and host cells comprising a nucleic acid of the invention; the
 CC recombinant production of a polypeptide of the invention; an antibody
 CC against a polypeptide of the invention; a method of detecting
 CC polynucleotides or polypeptides of the invention; and methods of
 CC identifying a compound which binds to a polypeptide of the invention. The
 CC invention further discloses methods of patenting, treating or
 CC ameliorating a medical condition; kits comprising polynucleotide probes
 CC and/or monoclonal antibodies for carrying out the methods of the
 CC invention; methods for the identification of compounds that modulate the
 CC expression or activity of the polynucleotide and/or polypeptide; and 767
 CC contig sequences corresponding to the cDNA sequences of the invention
 CC (ADC31861-ADC32627) and the polypeptides encoded by the contigs (ADC32628
 CC -ADC33394). The nucleic acids and polypeptides of the invention are
 CC useful in diagnostics, drug screening, forensics, gene mapping, in the
 CC identification of mutations responsible for genetic disorders or other
 CC traits, for assessing biodiversity, and in producing many other types of
 CC data and products dependent on DNA and amino acid sequences. They are
 CC also used for treating diseases such as Parkinson's disease, Alzheimer's
 CC disease and other neurodegenerative diseases, anaemia, platelet
 CC disorders, wounds, ulcers, osteoporosis, autoimmune diseases or
 CC cancer. The nucleic acids may also be used as hybridisation probes or
 CC primers, and in the recombinant production of a protein. The polypeptides
 CC are also useful in generating antibodies, as molecular weight markers,
 CC and as food supplements. The present sequence represents a specifically
 CC claimed human polypeptide sequence of the invention. Note: The sequence
 CC data for this patent did not form part of the printed specification, but
 CC was obtained in electronic format directly from WIPO at
 CC ftp.wipo.int/pub/published_pct_sequences.

XX Sequence 416 AA;

Query Match 1.5%; Score 8; DB 7; length 416;

Best Local Similarity 100.0%; Pred. No. 1.1e-02;

Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 299 QVELREEH 306

Db 20 QVELREEH 27

RESULT 40

AAB79356

ID AAB79356 standard; protein; 484 AA.

AC AAB79356;

DT 30-APR-2001 (first entry)

DE Corynebacterium glutamicum SMP protein sequence SEQ ID NO:228.

XX Corynebacterium glutamicum; carbon metabolism and energy production;

XX SMP protein; sugar metabolism and oxidative phosphorylation protein;

XX fine chemical production; organic acid; proteinogenic amino acid;

XX nonproteinogenic amino acid; purine base; pyrimidine base; nucleoside;

XX nucleotide; lipid; saturated fatty acid; unsaturated fatty acid; diol;

XX carbohydrate; aromatic compound; vitamin; cofactor; polypeptide; enzyme;

XX diagnosis; Corynebacterium diphtheriae; evolutionary study.

OS Corynebacterium glutamicum.

XX WO200100844-A2.

XX 04-JAN-2001.

XX 23-JUN-2000; 2000WG-IB000943.

XX 25-JUN-1999; 99US-0141031P.

XX 08-JUL-1999; 99DE-01031412.

XX 08-JUL-1999; 99DE-01031413.

PR 08-JUL-1999; 99DE-01031419.
 PR 08-JUL-1999; 99DE-01031420.
 PR 08-JUL-1999; 99DE-01031424.
 PR 08-JUL-1999; 99DE-01031428.
 PR 08-JUL-1999; 99DE-01031431.
 PR 08-JUL-1999; 99DE-01031433.
 PR 08-JUL-1999; 99DE-01031434.
 PR 08-JUL-1999; 99DE-01031510.
 PR 08-JUL-1999; 99DE-01031562.
 PR 08-JUL-1999; 99DE-01031634.
 PR 09-JUL-1999; 99DE-01032180.
 PR 09-JUL-1999; 99DE-01032227.
 PR 09-JUL-1999; 99DE-01032230.
 PR 09-JUL-1999; 99US-0143208P.
 PR 14-JUL-1999; 99DE-01032924.
 PR 14-JUL-1999; 99DE-01032973.
 PR 14-JUL-1999; 99DE-01033005.
 PR 27-AUG-1999; 99DE-01040765.
 PR 31-AUG-1999; 99US-0151572P.
 PR 03-SEP-1999; 99DE-01042076.
 PR 03-SEP-1999; 99DE-01042086.
 PR 03-SEP-1999; 99DE-01042087.
 PR 03-SEP-1999; 99DE-01042088.
 PR 03-SEP-1999; 99DE-01042095.
 PR 03-SEP-1999; 99DE-01042123.
 PR 03-SEP-1999; 99DE-01042125.

XX (BADI) BASF AG.

PA Pompejus M, Kroeger B, Schroeder H, Zelder O, Haberhauer G;

PI WPI: 2001-061975/07.

XX N-PSDB; AAF71473.

XX New isolated Corynebacterium glutamicum nucleic acid encoding a sugar
 PT metabolism and oxidative phosphorylation protein for production or
 PT modulation of production of fine chemicals e.g. amino acids,
 PT carbohydrates or enzymes.

Claim 20; Page 464-466; 1246pp; English.

CC AAF71360 to AAF71750 encode the Corynebacterium glutamicum sugar
 CC metabolism and oxidative phosphorylation (SMP) proteins given in AAB79243
 CC to AAB 79633 which are involved in carbon metabolism and energy
 CC production. The C. glutamicum SMP gene can be used in vectors (II) for
 CC expression in host cells and production or modulation of production of
 CC fine chemicals, such as, an organic acid, a proteinogenic or
 CC nonproteinogenic amino acid (preferred), a purine or pyrimidine base, a
 CC nucleoside, a nucleotide, a lipid, a saturated or unsaturated fatty acid,
 CC a diol, a carbohydrate, an aromatic compound, a vitamin, a cofactor, a
 CC polyketide, or an enzyme. The presence of (I) or SMP proteins (III)
 CC encoded by them are used for diagnosing the presence or activity of
 CC Corynebacterium diphtheriae in a subject. (I), (II), (III) or host cells
 CC containing them are used to map genomes of organisms related to C.
 CC glutamicum, identify and localise C. glutamicum sequences of interest, in
 CC evolutionary studies, in determining SMP protein regions required for
 CC function, in modulating SMP protein activity, in modulating the
 CC metabolism of sugars, and in modulating high-energy molecule production
 CC in a cell (i.e. ATP, NADPH)

XX Sequence 484 AA;

Query Match 1.5%; Score 8; DB 4; Length 484;

Best Local Similarity 100.0%; Pred. No. 1.3e+02;

Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 181 AELLEHR 188

Db 80 AELLEHR 87

RESULT 41

AAG92676	KW	hybridisation assay; genetic mapping; gene expression control; promoter; termination sequence.
ID AAG92676 standard; protein; 484 AA.	KW	
AC AAG92676;	XX	
XX	OS	Arabidopsis thaliana.
DT 26-SEP-2001 (first entry)	PN	EP1033405-A2.
XX	XX	06-SEP-2000.
XX	PD	
DE C glutamic protein fragment SEQ ID NO: 6430.	XX	25-FEB-2000; 2000EP-00301439.
XX	PF	
KW Corynebacterium; amino acid synthesis; vitamin; saccharide; organic acid synthesis.	XX	25-FEB-1999; 99US-0121825P.
XX	PR	05-MAR-1999; 99US-0123180P.
XX	PR	09-MAR-1999; 99US-0123548P.
OS Corynebacterium glutamicum.	PR	23-MAR-1999; 99US-0125788P.
XX	PR	25-MAR-1999; 99US-0126264P.
PN EP1108790-A2.	PR	29-MAR-1999; 99US-0125785P.
XX	PR	01-APR-1999; 99US-0127462P.
PD 20-JUN-2001.	PR	06-APR-1999; 99US-0128334P.
XX	PR	08-APR-1999; 99US-0128714P.
XX	PR	16-APR-1999; 99US-0129845P.
XX	PR	19-APR-1999; 99US-0130077P.
XX	PR	21-APR-1999; 99US-0130449P.
XX	PR	23-APR-1999; 99US-0130510P.
XX	PR	23-APR-1999; 99US-0130891P.
XX	PR	28-APR-1999; 99US-0131449P.
XX	PR	30-APR-1999; 99US-0132048P.
XX	PR	03-APR-1999; 99US-0132407P.
XX	PR	04-MAY-1999; 99US-0132484P.
XX	PR	05-MAY-1999; 99US-0132485P.
XX	PR	06-MAY-1999; 99US-0132486P.
XX	PR	06-MAY-1999; 99US-0132487P.
XX	PR	07-MAY-1999; 99US-0132863P.
XX	PR	11-MAY-1999; 99US-0134256P.
XX	PR	14-MAY-1999; 99US-0134218P.
XX	PR	14-MAY-1999; 99US-0134219P.
XX	PR	14-MAY-1999; 99US-0134221P.
XX	PR	14-MAY-1999; 99US-0134370P.
XX	PR	18-MAY-1999; 99US-0134768P.
XX	PR	19-MAY-1999; 99US-0134941P.
XX	PR	20-MAY-1999; 99US-0135124P.
XX	PR	21-MAY-1999; 99US-0135353P.
XX	PR	24-MAY-1999; 99US-0135629P.
XX	PR	25-MAY-1999; 99US-0136021P.
XX	PR	27-MAY-1999; 99US-0136392P.
XX	PR	28-MAY-1999; 99US-0136782P.
XX	PR	01-JUN-1999; 99US-0137222P.
XX	PR	03-JUN-1999; 99US-0137528P.
XX	PR	04-JUN-1999; 99US-0137502P.
XX	PR	07-JUN-1999; 99US-0137724P.
XX	PR	08-JUN-1999; 99US-0138094P.
XX	PR	10-JUN-1999; 99US-0138540P.
XX	PR	10-JUN-1999; 99US-0138847P.
XX	PR	14-JUN-1999; 99US-0139119P.
XX	PR	16-JUN-1999; 99US-0139452P.
XX	PR	16-JUN-1999; 99US-0139453P.
XX	PR	17-JUN-1999; 99US-0139492P.
XX	PR	18-JUN-1999; 99US-0139454P.
XX	PR	18-JUN-1999; 99US-0139455P.
XX	PR	18-JUN-1999; 99US-0139456P.
XX	PR	18-JUN-1999; 99US-0139457P.
XX	PR	18-JUN-1999; 99US-0139458P.
XX	PR	18-JUN-1999; 99US-0139459P.
XX	PR	18-JUN-1999; 99US-0139460P.
XX	PR	18-JUN-1999; 99US-0139461P.
XX	PR	18-JUN-1999; 99US-0139462P.
XX	PR	18-JUN-1999; 99US-0139463P.
XX	PR	18-JUN-1999; 99US-0139750P.
XX	PR	18-JUN-1999; 99US-0139763P.
XX	PR	21-JUN-1999; 99US-0139817P.
XX	PR	22-JUN-1999; 99US-0139899P.
XX	PR	23-JUN-1999; 99US-0140353P.
XX	PR	23-JUN-1999; 99US-0140354P.
XX	PR	

Query Match 1.5%; Score 8; DB 4; Length 484;
Best Local Similarity 100.0%; Pred. No. 1.3e+02;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 181 AELLLEHR 188
|||
Db 80 AELLLEHR 87

RESULT 42
AAG41279
ID AAG41279 standard; protein; 785 AA.
XX
AC AAG41279;
XX
DT 18-OCT-2000 (first entry)
XX
DE Arabidopsis thaliana protein fragment SEQ ID NO: 51338.
XX
XX Protein identification; signal transduction pathway; metabolic pathway;

```
PR 24-JUN-1999; 99US-0140695P.
PR 28-JUN-1999; 99US-0140823P.
PR 29-JUN-1999; 99US-0140991P.
PR 30-JUN-1999; 99US-0141287P.
PR 01-JUL-1999; 99US-0141842P.
PR 01-JUL-1999; 99US-0142154P.
PR 02-JUL-1999; 99US-0142055P.
PR 06-JUL-1999; 99US-0142390P.
PR 08-JUL-1999; 99US-0142803P.
PR 09-JUL-1999; 99US-0142920P.
PR 12-JUL-1999; 99US-0142977P.
PR 13-JUL-1999; 99US-0143542P.
PR 14-JUL-1999; 99US-0143824P.
PR 15-JUL-1999; 99US-0144005P.
PR 16-JUL-1999; 99US-0144085P.
PR 16-JUL-1999; 99US-0144086P.
PR 19-JUL-1999; 99US-0144325P.
PR 19-JUL-1999; 99US-0144331P.
PR 19-JUL-1999; 99US-0144332P.
PR 19-JUL-1999; 99US-0144333P.
PR 19-JUL-1999; 99US-0144334P.
PR 19-JUL-1999; 99US-0144335P.
PR 20-JUL-1999; 99US-0144352P.
PR 20-JUL-1999; 99US-0144632P.
PR 20-JUL-1999; 99US-0144884P.
PR 21-JUL-1999; 99US-0144814P.
PR 21-JUL-1999; 99US-0145086P.
PR 21-JUL-1999; 99US-0145088P.
PR 21-JUL-1999; 99US-0145085P.
PR 22-JUL-1999; 99US-0145087P.
PR 22-JUL-1999; 99US-0145088P.
PR 22-JUL-1999; 99US-0145192P.
PR 23-JUL-1999; 99US-0145145P.
PR 23-JUL-1999; 99US-0145218P.
PR 23-JUL-1999; 99US-0145224P.
PR 26-JUL-1999; 99US-0145276P.
PR 27-JUL-1999; 99US-0145913P.
PR 27-JUL-1999; 99US-0145918P.
PR 27-JUL-1999; 99US-0145919P.
PR 28-JUL-1999; 99US-0145951P.
PR 02-AUG-1999; 99US-0146386P.
PR 02-AUG-1999; 99US-0146388P.
PR 02-AUG-1999; 99US-0146389P.
PR 03-AUG-1999; 99US-0147038P.
PR 04-AUG-1999; 99US-0147204P.
PR 05-AUG-1999; 99US-0147302P.
PR 05-AUG-1999; 99US-0147192P.
PR 05-AUG-1999; 99US-0147260P.
PR 06-AUG-1999; 99US-0147403P.
PR 06-AUG-1999; 99US-0147416P.
PR 09-AUG-1999; 99US-0147935P.
PR 09-AUG-1999; 99US-0148171P.
PR 10-AUG-1999; 99US-0148119P.
PR 11-AUG-1999; 99US-0148341P.
PR 12-AUG-1999; 99US-0148565P.
PR 13-AUG-1999; 99US-0148684P.
PR 15-AUG-1999; 99US-0149368P.
PR 17-AUG-1999; 99US-0149175P.
PR 18-AUG-1999; 99US-0149426P.
PR 20-AUG-1999; 99US-0149722P.
PR 20-AUG-1999; 99US-0149723P.
PR 20-AUG-1999; 99US-0149929P.
PR 23-AUG-1999; 99US-0149902P.
PR 23-AUG-1999; 99US-0149930P.
PR 25-AUG-1999; 99US-0150566P.
PR 26-AUG-1999; 99US-0150884P.
PR 26-AUG-1999; 99US-0151065P.
PR 27-AUG-1999; 99US-0151066P.
PR 27-AUG-1999; 99US-0151080P.
PR 30-AUG-1999; 99US-0151303P.
PR 31-AUG-1999; 99US-0151438P.
PR 01-SEP-1999; 99US-0151930P.

PR 07-SEP-1999; 99US-0152363P.
PR 10-SEP-1999; 99US-0153070P.
PR 13-SEP-1999; 99US-0153758P.
PR 15-SEP-1999; 99US-0154018P.
PR 16-SEP-1999; 99US-0154039P.
PR 20-SEP-1999; 99US-0154779P.
PR 22-SEP-1999; 99US-0155139P.
PR 23-SEP-1999; 99US-0155486P.
PR 24-SEP-1999; 99US-0155659P.
PR 28-SEP-1999; 99US-0156458P.
PR 29-SEP-1999; 99US-0156596P.
PR 04-OCT-1999; 99US-0157117P.
PR 05-OCT-1999; 99US-0157533P.
PR 06-OCT-1999; 99US-0157865P.
PR 07-OCT-1999; 99US-0158029P.
PR 08-OCT-1999; 99US-0158232P.
PR 12-OCT-1999; 99US-0158369P.
PR 13-OCT-1999; 99US-0159293P.
PR 13-OCT-1999; 99US-0159294P.
PR 13-OCT-1999; 99US-0159295P.
PR 14-OCT-1999; 99US-0159330P.
PR 14-OCT-1999; 99US-0159331P.
PR 14-OCT-1999; 99US-0159637P.
PR 14-OCT-1999; 99US-0159638P.
PR 18-OCT-1999; 99US-0159584P.
PR 21-OCT-1999; 99US-0160741P.
PR 21-OCT-1999; 99US-0160767P.
PR 21-OCT-1999; 99US-0160768P.
PR 21-OCT-1999; 99US-0160770P.
PR 21-OCT-1999; 99US-0160814P.
PR 21-OCT-1999; 99US-0160815P.
PR 22-OCT-1999; 99US-0160980P.
PR 22-OCT-1999; 99US-0160981P.
PR 22-OCT-1999; 99US-0160989P.
PR 22-OCT-1999; 99US-0161404P.
PR 25-OCT-1999; 99US-0161405P.
PR 25-OCT-1999; 99US-0161406P.
PR 26-OCT-1999; 99US-0161359P.
PR 26-OCT-1999; 99US-0161360P.
PR 26-OCT-1999; 99US-0161361P.
PR 28-OCT-1999; 99US-0161920P.
PR 28-OCT-1999; 99US-0161992P.
PR 28-OCT-1999; 99US-0161993P.
PR 29-OCT-1999; 99US-0162142P.

Query Match 1.5%; Score 8; DB 3; Length 785;
Best Local Similarity 100.0%; Pred. No. 1.9e+02; Mismatches 0; Indels 0; Gaps 0;
Matches 8; Conservative 0;

QY 127 EEIRQSDE 134
Db 216 EEIRQSDE 223
|||||
|||||

RESULT 43
AAG41278
ID AAG41278 standard; protein; 823 AA.
XX
AC AAG41278;
XX
DT 18-OCT-2000 (first entry)
XX
DE Arabidopsis thaliana protein fragment SEQ ID NO: 51337.
XX
KW Protein identification; signal transduction pathway; metabolic pathway;
hybridisation assay; genetic mapping; gene expression control; promoter;
termination sequence.
XX
OS Arabidopsis thaliana.
XX
PN EP1033405-A2.
XX
```

PD 06-SEP-2000. 99US-0142390P.
XX 08-JUL-1999; 99US-0142803P.
PF 09-JUL-1999; 99US-0142920P.
XX 12-JUL-1999; 99US-0142977P.
PR 13-JUL-1999; 99US-0143542P.
PR 14-JUL-1999; 99US-0143624P.
PR 15-JUL-1999; 99US-0144005P.
PR 16-JUL-1999; 99US-0144085P.
PR 16-JUL-1999; 99US-0144086P.
PR 19-JUL-1999; 99US-0144325P.
PR 19-JUL-1999; 99US-0144331P.
PR 19-JUL-1999; 99US-0144332P.
PR 19-JUL-1999; 99US-0144333P.
PR 19-JUL-1999; 99US-0144334P.
PR 19-JUL-1999; 99US-0144335P.
PR 20-JUL-1999; 99US-0144352P.
PR 20-JUL-1999; 99US-0144632P.
PR 20-JUL-1999; 99US-0144684P.
PR 21-JUL-1999; 99US-0144814P.
PR 21-JUL-1999; 99US-0145086P.
PR 21-JUL-1999; 99US-0145088P.
PR 22-JUL-1999; 99US-0145085P.
PR 22-JUL-1999; 99US-0145087P.
PR 22-JUL-1999; 99US-0145089P.
PR 22-JUL-1999; 99US-0145192P.
PR 23-JUL-1999; 99US-0145145P.
PR 23-JUL-1999; 99US-0145218P.
PR 23-JUL-1999; 99US-0145224P.
PR 26-JUL-1999; 99US-0145276P.
PR 27-JUL-1999; 99US-0145913P.
PR 27-JUL-1999; 99US-0145918P.
PR 28-JUL-1999; 99US-0145919P.
PR 28-JUL-1999; 99US-0145951P.
PR 02-AUG-1999; 99US-0146386P.
PR 02-AUG-1999; 99US-0146388P.
PR 02-AUG-1999; 99US-0146389P.
PR 03-AUG-1999; 99US-0147038P.
PR 04-AUG-1999; 99US-0147204P.
PR 04-AUG-1999; 99US-0147302P.
PR 05-AUG-1999; 99US-0147192P.
PR 05-AUG-1999; 99US-0147260P.
PR 06-AUG-1999; 99US-0147303P.
PR 06-AUG-1999; 99US-0147416P.
PR 09-AUG-1999; 99US-0147493P.
PR 09-AUG-1999; 99US-0147935P.
PR 10-AUG-1999; 99US-0148171P.
PR 11-AUG-1999; 99US-0148319P.
PR 12-AUG-1999; 99US-0148341P.
PR 13-AUG-1999; 99US-0148565P.
PR 13-AUG-1999; 99US-0148684P.
PR 16-AUG-1999; 99US-0149368P.
PR 17-AUG-1999; 99US-0149175P.
PR 18-AUG-1999; 99US-0149426P.
PR 20-AUG-1999; 99US-0149722P.
PR 20-AUG-1999; 99US-0149723P.
PR 20-AUG-1999; 99US-0149929P.
PR 23-AUG-1999; 99US-0149902P.
PR 25-AUG-1999; 99US-0149930P.
PR 26-AUG-1999; 99US-0150566P.
PR 26-AUG-1999; 99US-0150884P.
PR 27-AUG-1999; 99US-0151065P.
PR 27-AUG-1999; 99US-0151066P.
PR 27-AUG-1999; 99US-0151080P.
PR 30-AUG-1999; 99US-0151303P.
PR 31-AUG-1999; 99US-0151438P.
PR 01-SEP-1999; 99US-0151930P.
PR 07-SEP-1999; 99US-0152363P.
PR 10-SEP-1999; 99US-0153070P.
PR 13-SEP-1999; 99US-0153758P.
PR 15-SEP-1999; 99US-0154018P.
PR 16-SEP-1999; 99US-0154039P.
PR 20-SEP-1999; 99US-0154779P.
PR 22-SEP-1999; 99US-0155139P.

PR 23-SEP-1999; 99US-0155486P.
 PR 24-SEP-1999; 99US-0155659P.
 PR 28-SEP-1999; 99US-0156458P.
 PR 29-SEP-1999; 99US-0155596P.
 PR 04-OCT-1999; 99US-0157117P.
 PR 05-OCT-1999; 99US-0157753P.
 PR 06-OCT-1999; 99US-0157865P.
 PR 07-OCT-1999; 99US-0158029P.
 PR 08-OCT-1999; 99US-0158232P.
 PR 12-OCT-1999; 99US-0158369P.
 PR 13-OCT-1999; 99US-0158293P.
 PR 13-OCT-1999; 99US-0158294P.
 PR 13-OCT-1999; 99US-0159295P.
 PR 14-OCT-1999; 99US-0159329P.
 PR 14-OCT-1999; 99US-0159330P.
 PR 14-OCT-1999; 99US-0159331P.
 PR 14-OCT-1999; 99US-0159637P.
 PR 14-OCT-1999; 99US-0159638P.
 PR 18-OCT-1999; 99US-0159584P.
 PR 21-OCT-1999; 99US-0160741P.
 PR 21-OCT-1999; 99US-0160767P.
 PR 21-OCT-1999; 99US-0160768P.
 PR 21-OCT-1999; 99US-0160770P.
 PR 21-OCT-1999; 99US-0160814P.
 PR 21-OCT-1999; 99US-0160815P.
 PR 22-OCT-1999; 99US-0160980P.
 PR 22-OCT-1999; 99US-0160981P.
 PR 22-OCT-1999; 99US-0160989P.
 PR 25-OCT-1999; 99US-0161404P.
 PR 25-OCT-1999; 99US-0161405P.
 PR 25-OCT-1999; 99US-0161406P.
 PR 26-OCT-1999; 99US-0161359P.
 PR 26-OCT-1999; 99US-0161360P.
 PR 26-OCT-1999; 99US-0161361P.
 PR 28-OCT-1999; 99US-0161920P.
 PR 28-OCT-1999; 99US-0161992P.
 PR 28-OCT-1999; 99US-0161993P.
 PR 29-OCT-1999; 99US-0162142P.
 PR 29-OCT-1999; 99US-0162143P.

Query Match 1.5%; Score 8; DB 3; Length 823;

Best Local Similarity 100.0%; Pred. No. 2e+02; 0; Indels 0; Gaps 0;

Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 127 EIRQSD 134
 DB 254 EIRQSD 261

RESULT 44
 AAG41277
 ID AAG41277 standard; protein; 836 AA.

XX AC AAG41277;
 XX DT 18-OCT-2000 (first entry)

XX DE Arabidopsis thaliana protein fragment SEQ ID NO: 51336.

XX KW Protein identification; signal transduction pathway; metabolic pathway;
 XX KW hybridisation assay; genetic mapping; gene expression control; promoter;
 XX KW termination sequence.

XX OS Arabidopsis thaliana.
 XX PN EP1033405-A2.

XX PD 06-SEP-2000.

XX XX 25-FEB-2000; 2000EP-00301439.

XX XX 25-FEB-1999; 99US-0121825P.

XX XX 05-MAR-1999; 99US-0123180P.

XX XX 09-MAR-1999; 99US-0123548P.

PR 23-MAR-1999; 99US-0125788P.
 PR 25-MAR-1999; 99US-0126264P.
 PR 29-MAR-1999; 99US-0126785P.
 PR 01-APR-1999; 99US-0127462P.
 PR 06-APR-1999; 99US-0128234P.
 PR 08-APR-1999; 99US-0128714P.
 PR 16-APR-1999; 99US-0129845P.
 PR 19-APR-1999; 99US-0130077P.
 PR 21-APR-1999; 99US-0130449P.
 PR 23-APR-1999; 99US-0130510P.
 PR 28-APR-1999; 99US-0131449P.
 PR 30-APR-1999; 99US-0132048P.
 PR 04-MAY-1999; 99US-0132407P.
 PR 05-MAY-1999; 99US-0132484P.
 PR 06-MAY-1999; 99US-0132485P.
 PR 06-MAY-1999; 99US-0132486P.
 PR 07-MAY-1999; 99US-0132487P.
 PR 11-MAY-1999; 99US-0132863P.
 PR 14-MAY-1999; 99US-0134218P.
 PR 14-MAY-1999; 99US-0134219P.
 PR 14-MAY-1999; 99US-0134221P.
 PR 14-MAY-1999; 99US-0134370P.
 PR 18-MAY-1999; 99US-0134768P.
 PR 19-MAY-1999; 99US-0134941P.
 PR 20-MAY-1999; 99US-0135124P.
 PR 21-MAY-1999; 99US-0135353P.
 PR 24-MAY-1999; 99US-0135629P.
 PR 25-MAY-1999; 99US-0136021P.
 PR 27-MAY-1999; 99US-0136392P.
 PR 28-MAY-1999; 99US-0136782P.
 PR 01-JUN-1999; 99US-0137222P.
 PR 03-JUN-1999; 99US-0137528P.
 PR 04-JUN-1999; 99US-0137502P.
 PR 07-JUN-1999; 99US-0137724P.
 PR 08-JUN-1999; 99US-0138094P.
 PR 10-JUN-1999; 99US-0138540P.
 PR 10-JUN-1999; 99US-0138847P.
 PR 14-JUN-1999; 99US-0139119P.
 PR 16-JUN-1999; 99US-0139452P.
 PR 16-JUN-1999; 99US-0139453P.
 PR 17-JUN-1999; 99US-0139492P.
 PR 18-JUN-1999; 99US-0139454P.
 PR 18-JUN-1999; 99US-0139455P.
 PR 18-JUN-1999; 99US-0139456P.
 PR 18-JUN-1999; 99US-0139457P.
 PR 18-JUN-1999; 99US-0139458P.
 PR 18-JUN-1999; 99US-0139459P.
 PR 18-JUN-1999; 99US-0139460P.
 PR 18-JUN-1999; 99US-0139461P.
 PR 18-JUN-1999; 99US-0139462P.
 PR 18-JUN-1999; 99US-0139463P.
 PR 18-JUN-1999; 99US-0139750P.
 PR 18-JUN-1999; 99US-0139763P.
 PR 21-JUN-1999; 99US-0139817P.
 PR 22-JUN-1999; 99US-0139899P.
 PR 23-JUN-1999; 99US-0140353P.
 PR 23-JUN-1999; 99US-0140354P.
 PR 24-JUN-1999; 99US-0140695P.
 PR 28-JUN-1999; 99US-0140823P.
 PR 29-JUN-1999; 99US-0140991P.
 PR 30-JUN-1999; 99US-0141287P.
 PR 01-JUL-1999; 99US-0141842P.
 PR 01-JUL-1999; 99US-0142154P.
 PR 02-JUL-1999; 99US-0142055P.
 PR 06-JUL-1999; 99US-0142390P.
 PR 08-JUL-1999; 99US-0142803P.
 PR 09-JUL-1999; 99US-0142920P.
 PR 12-JUL-1999; 99US-0142977P.
 PR 13-JUL-1999; 99US-0143542P.
 PR 14-JUL-1999; 99US-0143624P.
 PR 15-JUL-1999; 99US-0144005P.

Query Match	Score 8;	DB 3;	Length 836;
Best Local Similarity	100.0%;	Pred. No. 2e+02;	
Matches	8;	Conservative	0; Mismatches 0; Indels 0; Gaps 0;
QY	127	BEIROSDE 134	
DB	267	BEIROSDE 274	
<p>RESULT 45</p> <p>ADB95084</p> <p>ID ADB95084 standard; protein; 952 AA.</p> <p>XX</p> <p>AC ADB95084;</p> <p>XX</p> <p>DT 04-DEC-2003 (first entry)</p> <p>XX</p> <p>DE A. thaliana protein 62837 #SEQ ID 82.</p> <p>XX</p> <p>KW Plant; herbicide; weed; crop field; growth; development.</p> <p>XX</p> <p>OS Arabidopsis thaliana.</p> <p>XX</p> <p>PN WC2003008440-A2.</p> <p>XX</p> <p>PD 30-JAN-2003.</p> <p>XX</p> <p>PF 16-JUL-2002; 2002WO-EP007929.</p> <p>XX</p> <p>PR 16-JUL-2001; 2001US-0305806P.</p> <p>PR 20-FEB-2002; 2002US-0358416P.</p> <p>XX</p> <p>PA (SYGN) SYNGENTA PARTICIPATIONS AG.</p> <p>XX</p> <p>PI Levin JZ, Patton DA, Mcelver JA, Budziszewski CJ, Zhou Q, Aux GW;</p> <p>PI Tossberg J, Wegrich Glover L, Ashby CS, Thomas CR, Madhaven E;</p> <p>PI Lewis S, Dunn J, Cates E, Law MD;</p> <p>XX</p> <p>DR WPI; 2003-229557/22.</p> <p>DR N-PSDB; ADB95083.</p> <p>XX</p>			

PT Identifying an herbicidal compound, useful for controlling undesirable
PT vegetation, comprises combining a polypeptide with a compound to be
PT tested for the ability to bind to the polypeptide or inhibit the activity
XX of the polypeptide.

PS Claim 4; SEQ ID NO 82; 273pp; English.

XX The invention relates to a method for identifying a herbicidal compound.
CC The method of the invention comprises combining a polypeptide having at
CC least 90% identical to any one of 48 69-1008 residue amino acid sequences
CC (designated as PI-P48), given in the specification, with a compound to be
CC tested for the ability to bind to the polypeptide or inhibit the activity
CC of the polypeptide, under conditions conducive to binding or inhibiting,
CC respectively. Also disclosed is a method for killing or inhibiting the
CC growth or viability of a plant by applying to the plant the herbicidal
CC compound identified by the novel method, a chimeric construct comprising
CC a promoter operatively linked to the nucleic acid molecule, a recombinant
CC vector comprising the chimeric construct and a host cell comprising the
CC nucleic acid molecule. The method and polypeptides are useful in
CC screening assays to identify compounds that interact or inhibit the
CC polypeptides, thus as potential herbicides to control undesirable
CC vegetation such as weeds in crop fields. Nucleic acid molecules (odd
CC numbers between ADB95003 and ADB95097) isolated from Arabidopsis thaliana
CC comprising nucleotide sequences that encode proteins (even numbers
CC between ADB95004-ADB95098) are essential for plant growth and
CC development.

XX Sequence 952 AA;

Query Match 1.5%; Score 8; DB 7; Length 952;
Best Local Similarity 100.0%; Pred. No. 2.3e+02;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 127 BEIRQSD 134
|||
DB 381 BEIRQSD 388

RESULT 46

ID ABOG8332 standard; protein; 1045 AA.

XX ABOG8332;

DT 13-FEB-2002 (first entry)

DE Novel human diagnostic protein #8323.

KW Human; chromosome mapping; gene mapping; gene therapy; forensic;
KW food supplement; medical imaging; diagnostic; genetic disorder.

XX Homo sapiens.

FN WO200175067-A2.

XX 11-OCT-2001.

PF 30-MAR-2001; 2001WO-US008631.

PR 31-MAR-2000; 2000US-00540217.

PR 23-AUG-2000; 2000US-00649167.

XX (HYSE-) HYSEQ INC.

XX Drmanac RT, Liu C, Tang YT;

XX WPI; 2001-639362/73.

DR N-PSDB; AAS72519.

XX New isolated polynucleotide and encoded polypeptides, useful in
PT diagnostics, forensics, gene mapping, identification of mutations
PT responsible for genetic disorders or other traits and to assess
PT biodiversity.

XX Claim 20; SEQ ID NO 38691; 103pp; English.

XX The invention relates to isolated polynucleotide (I) and polypeptide (II)
CC sequences. (I) is useful as hybridisation probes, polymerase chain
CC reaction (PCR) primers, oligomers, and for chromosome and gene mapping,
CC and in recombinant production of (II). The polynucleotides are also used
CC in diagnostics as expressed sequence tags for identifying expressed
CC genes. (I) is useful in gene therapy techniques to restore normal
CC activity of (II) or to treat disease states involving (II). (II) is
CC useful for generating antibodies against it, detecting or quantitating a
CC polypeptide in tissue, as molecular weight markers and as a food
CC supplement. (II) and its binding partners are useful in medical imaging
CC of sites expressing (II). (I) and (II) are useful for treating disorders
CC involving aberrant protein expression or biological activities in
CC polypeptide and polynucleotide sequences have applications in
CC diagnostics, forensics, gene mapping, identification of mutations
CC responsible for genetic disorders or other traits to assess biodiversity
CC and to produce other types of data and products dependent on DNA and
CC amino acid sequences. ABOG0010-ABG30377 represent novel human diagnostic
CC amino acid sequences. The sequence data for this
CC patent did not appear in the printed specification, but was obtained in
CC electronic format directly from WIPO at
CC ftp.wipo.int/pub/published_pct_sequences

XX Sequence 1045 AA;

Query Match 1.5%; Score 8; DB 4; Length 1045;

Best Local Similarity 100.0%; Pred. No. 2.4e+02;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 513 STRASGQT 520
|||||
DB 783 STRASGQT 790

RESULT 47

AAB40945

ID AAB40945 standard; protein; 1532 AA.

XX AAB40945;

DT 08-FEB-2001 (first entry)

DE Human ORFX ORF709 polypeptide sequence SEQ ID NO:1418.

KW Human; open reading frame; ORFX; detection; cytostatic; hepatotropic;
KW vulnery; antipsoriatic; antiparkinsonian; nootropic; neuroprotective;
KW anticonvulsant; osteopathic; antiarthritic; immunosuppressant; cardiant;
KW immunostimulant; thrombolytic; coagulant; vasotropic; antidiabetic;
KW hypotensive; dermatological; immunosuppressive; antinflammatory;
KW antiviral; antibacterial; antifungal; antirheumatic; antithyroid;
KW antianaemic; gene therapy; cancer; proliferative disorder; hypertension;
KW neurodegenerative disorder; osteoarthritis; graft vs host disease;
KW cardiovascular disease; diabetes mellitus; hypothyroidism; SCID; AIDS;
KW cholesterol ester storage; systemic lupus erythematosus; infection;
KW severe combined immunodeficiency; malaria; autolimmune disorder; asthma;
KW allergy; aplastic anaemia; nocturnal haemoglobinuria; burn; wound;
KW bone damage; cartilage damage; antinflammatory disease; coagulation;
KW thrombosis; contraceptive.

XX Homo sapiens.

FN WO2000058473-A2.

XX 05-OCT-2000.

XX 31-MAR-2000; 2000WO-US008621.

XX 31-MAR-1999; 99US-0127607P.

PR 02-APR-1999; 99US-0127636P.

PR 05-APR-1999; 99US-0127728P.

PR 30-MAR-2000; 2000US-00540763.

XX PA (CURA-) CURAGEN CORP.
XX PI Shimkets RA, Leach M;
XX PT WPI; 2000-602362/57.
XX DR N-PSDB; AAC75154.
XX PS
XX PT Novel nucleic acids and peptides derived from open reading frame X,
XX PT useful for treating e.g. cancers, proliferative disorders,
XX PT neurodegenerative disorders and cardiovascular disease.
XX PS
XX PS Claim 11; Page 1203-1206; 5507pp; English.
XX CC AAC74446 to AAC77606 encode the proteins given in AAB40237 to AAB43397,
XX CC which represent the human ORFX open reading frames 1 to 3161. The ORFX
XX CC sequences have activities such as: cytostatic; hepatotropic; vulnery;
XX CC antiproliferative; antipapillomatous; neuroprotective; osteopathic;
XX CC anticonvulsant; antiarthritic; immunosuppressant; immunostimulant;
XX CC cardiant; thrombolytic; coagulant; vasotrophic; antidiabetic; hypotensive;
XX CC dermatological; immunosuppressive; antineoplastic; antibacterial;
XX CC antiviral; antifungal; antirheumatic; antithyroid; and antianemic. The
XX CC sequences can be used for determining the presence of or predisposition
XX CC to, or preventing or treating pathological conditions associated with an
XX CC ORFX-associated disorder. The nucleic acids can be used to express ORFX
XX CC proteins in gene therapy vectors. The proteins and nucleic acids may be
XX CC used to treat cancers, proliferative disorders, neurodegenerative
XX CC disorders, osteoarthritis, graft vs host disease, cardiovascular disease,
XX CC diabetes mellitus, hypertension, hypothyroidism, cholesterol ester
XX CC storage, systemic lupus erythematosus, severe combined immunodeficiency
XX CC (SCID), AIDS, viral, bacterial or fungal infection, malaria, autoimmune
XX CC disorders, asthma, allergies, aplastic anaemia, burns, wounds, bone and
XX CC cartilage damage, nocturnal haemoglobinuria, antiinflammatory disease; to
XX CC enhance coagulation; to inhibit thrombosis; and as a contraceptive
XX SQ Sequence 1532 AA;

Query Match 1.5%; Score 8; DB 3; Length 1532;
Best Local Similarity 100.0%; Pred. No. 3.4e+02;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 513 STRASGQT 520
DB 783 STRASGQT 790

RESULT 48
AAB82820
ID AAB82820 standard; peptide; 7 AA.
XX AC AAB82820;
XX DT 12-NOV-2001 (first entry)
XX DE Human low density lipoprotein binding protein 3 (LBP-3) peptide.
XX KW Low density lipoprotein binding protein 3; LBP-3; LDL; human;
XX KW atherosclerosis; antiarteriosclerotic; therapy; diagnosis; vaccine.
XX OS Homo sapiens.
XX PI WO200164874-A2.
XX PD 07-SEP-2001.
XX PF 28-FEB-2001; 2001WO-US006356.
XX PR 02-MAR-2000; 2000US-00517849.
XX PR 14-JUL-2000; 2000US-00616289.
XX PA (BOST-) BOSTON HEART FOUND INC.
XX PI Lees AM, Lees RS, Law SW, Arjona AA;

WPI; 2001-565505/53.
New isolated low density lipoprotein binding polypeptide for treating,
diagnosing and/or identifying therapeutic agents for atherosclerosis.
Claim 14(e); Page; 143pp; English.
The present sequence is that of a peptide comprising amino acid residues
59-75 of novel human low density lipoprotein binding protein 3 (LBP-3), of
the invention that are capable of binding to native and methylated low
density lipoproteins (LDLs). Also claimed are biologically active
fragments and analogues of LBPs, polynucleotides encoding LBPs, as well
as expression vectors, cells and methods of producing the LBPs.
Polypeptides having the present amino acid sequence are among those
claimed. Methods of determining if an animal is at risk for
atherosclerosis, methods for evaluating an agent for use in treating
atherosclerosis, and methods for treating a cell having an abnormality in
structure or metabolism of LBP are also claimed, as are pharmaceutical
compositions comprising an LBP polypeptide or nucleic acid, and vaccine
compositions. Note: the present sequence is not shown in the
specification but is derived from the human LBP-3 sequence given in
figure 8A (see AAB82808)
Sequence 7 AA;

Query Match 1.3%; Score 7; DB 4; Length 7;
Best Local Similarity 100.0%; Pred. No. 1.4e+06;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 53 RDVSEEL 59
DB 1 RDVSEEL 7

RESULT 49
AAB50054
ID AAB50054 standard; peptide; 15 AA.
XX AC AAB50054;
XX DT 26-JUL-2002 (first entry)
XX DE Proline-enriched gamma-carboxyl glutamate-protein 194.05 N-terminal.
XX KW Proline-enriched gamma-carboxyl glutamate-protein 194.05; cancer;
XX KW inflammation.
XX OS Unidentified.
XX PN CN1331106-A.
XX PD 16-JAN-2002.
XX PF 30-JUN-2000; 2000CN-00116909.
XX PR 30-JUN-2000; 2000CN-00116909.
XX PA (BODE-) BODE GENE DEV CO LTD SHANGHAI.
XX PI Mao Y, Xie Y;
XX PD 2002-316393/36.
XX PT Proline-enriched gamma-carboxyl glutamate-protein 194.05 polypeptide and
XX PT its encoding polynucleotide, for treating e.g. cancer and inflammation.
XX PS Example 5; Page 20 (Disclosure); 37pp; Chinese.
XX CC This invention describes a novel proline-enriched gamma-carboxyl
XX CC glutamate-protein 194.05 polypeptide and its encoding nucleic acid. The
XX CC polynucleotide, polypeptide and its antagonist are useful for treating

XX PA (CURA-) CURAGEN CORP.
XX PI Shimkets RA, Leach M;
XX PT WPI; 2000-602362/57.
XX DR N-PSDB; AAC75154.
XX PS
XX PT Novel nucleic acids and peptides derived from open reading frame X,
XX PT useful for treating e.g. cancers, proliferative disorders,
XX PT neurodegenerative disorders and cardiovascular disease.
XX PS
XX PS Claim 11; Page 1203-1206; 5507pp; English.
XX CC AAC74446 to AAC77606 encode the proteins given in AAB40237 to AAB43397,
XX CC which represent the human ORFX open reading frames 1 to 3161. The ORFX
XX CC sequences have activities such as: cytostatic; hepatotropic; vulnery;
XX CC antiproliferative; antipapillomatous; neuroprotective; osteopathic;
XX CC anticonvulsant; antiarthritic; immunosuppressant; immunostimulant;
XX CC cardiant; thrombolytic; coagulant; vasotrophic; antidiabetic; hypotensive;
XX CC dermatological; immunosuppressive; antineoplastic; antibacterial;
XX CC antiviral; antifungal; antirheumatic; antithyroid; and antianemic. The
XX CC sequences can be used for determining the presence of or predisposition
XX CC to, or preventing or treating pathological conditions associated with an
XX CC ORFX-associated disorder. The nucleic acids can be used to express ORFX
XX CC proteins in gene therapy vectors. The proteins and nucleic acids may be
XX CC used to treat cancers, proliferative disorders, neurodegenerative
XX CC disorders, osteoarthritis, graft vs host disease, cardiovascular disease,
XX CC diabetes mellitus, hypertension, hypothyroidism, cholesterol ester
XX CC storage, systemic lupus erythematosus, severe combined immunodeficiency
XX CC (SCID), AIDS, viral, bacterial or fungal infection, malaria, autoimmune
XX CC disorders, asthma, allergies, aplastic anaemia, burns, wounds, bone and
XX CC cartilage damage, nocturnal haemoglobinuria, antiinflammatory disease; to
XX CC enhance coagulation; to inhibit thrombosis; and as a contraceptive
XX SQ Sequence 1532 AA;

Query Match 1.5%; Score 8; DB 3; Length 1532;
Best Local Similarity 100.0%; Pred. No. 3.4e+02;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 513 STRASGQT 520
DB 783 STRASGQT 790

RESULT 48
AAB82820
ID AAB82820 standard; peptide; 7 AA.
XX AC AAB82820;
XX DT 12-NOV-2001 (first entry)
XX DE Human low density lipoprotein binding protein 3 (LBP-3) peptide.
XX KW Low density lipoprotein binding protein 3; LBP-3; LDL; human;
XX KW atherosclerosis; antiarteriosclerotic; therapy; diagnosis; vaccine.
XX OS Homo sapiens.
XX PI WO200164874-A2.
XX PD 07-SEP-2001.
XX PF 28-FEB-2001; 2001WO-US006356.
XX PR 02-MAR-2000; 2000US-00517849.
XX PR 14-JUL-2000; 2000US-00616289.
XX PA (BOST-) BOSTON HEART FOUND INC.
XX PI Lees AM, Lees RS, Law SW, Arjona AA;

CC e.g. cancer and inflammation. This sequence represents the proline-
CC enriched gamma-carboxyl glutamate protein 194.05 N-terminal fragment
CC described in the method of the invention
XX
XX Sequence 15 AA;
SQ

Query Match 1.3%; Score 7; DB 5; Length 15;
Best Local Similarity 100.0%; Pred. No. 65;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 5 GQPEAGP 11
DB 6 GQPEAGP 12

RESULT 50
ADD24103
ID ADD24103 standard; peptide; 15 AA.
XX AC ADD24103;
XX DT 15-JAN-2004 (first entry)
XX DE Breast cancer membrane protein (BCMP) peptide.
XX KW breast cancer; screening; diagnosis; breast cancer therapy;
KW breast cancer membrane protein; BCMP; cytostatic; vaccine; human.
XX OS Homo sapiens.
XX PN WO2003087831-A2.
XX PD 23-OCT-2003.
XX PF 10-APR-2003; 2003WO-GB001559.
XX PR 11-APR-2002; 2002GB-00008331.
XX PA (OXFO-) OXFORD GLYCOSCIENCES UK LTD.
XX PI Hudson LJ, Stamps AC, Terrett JA;
XX DR WPI; 2003-045381/78.
XX
XX Screening, diagnosing and/or treating breast cancer by detecting a change
XX in expression or activity of a breast cancer membrane protein (BCMP)
XX polypeptide or encoding nucleic acid molecule.
XX
XX Example; Page 75; B1pp; English.
XX
XX The present invention describes a method of screening for and/or
XX diagnosing breast cancer in a subject, and/or monitoring the
XX effectiveness of breast cancer therapy. The method comprises detecting
XX and/or quantifying in a biological sample obtained from the subject a
XX breast cancer membrane protein (BCMP) polypeptide and a nucleic acid
XX molecule. Also described: (1) an antibody, its functionally-active
XX fragment, derivative or analogue, that specifically binds to one or more
XX of the BCMP polypeptide; (2) a diagnostic kit comprising a capture
XX reagent specific for an BCMP polypeptide, reagents and instructions for
XX use; (3) a method for screening for anti-breast cancer agents that
XX interact with the BCMP polypeptide, comprising contacting the polypeptide
XX with a candidate agent, and determining whether or not the candidate
XX agent interacts with the polypeptide; (4) a method for screening for anti-
XX breast cancer agents that modulate the expression or activity of an BCMP
XX polypeptide or the nucleic acid molecule cited above, comprising
XX comparing the expression or activity of the polypeptide or nucleic acid
XX molecule, in the presence and absence of a candidate agent or in the
XX presence of a control agent, and determining whether the candidate agent
XX causes the expression or activity of the polypeptide or nucleic acid
XX molecule to change; and (5) an agent identified by the method of (3) or
XX (4), which interacts with the polypeptide or causes the expression or
XX activity of the polypeptide, or the expression of the nucleic acid
XX molecule to change. BCMPs have cytostatic activities, and can be used in

CC vaccines. The BCMP polypeptide, nucleic acid molecule, antibody, agent or
CC their derivatives, are useful in the manufacture of a medicament for the
CC treatment of breast cancer, where the composition is a vaccine. The
CC present sequence represents a BCMP peptide which is used in the
CC exemplification of the present invention. N.B. The present sequence is
CC designated as SEQ ID NO:1054 on page 75, but does not correspond with the
CC previously given SEQ ID NO:1054 in the specification.
XX
XX Sequence 15 AA;
SQ

Query Match 1.3%; Score 7; DB 7; Length 15;
Best Local Similarity 100.0%; Pred. No. 65;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 473 SAGGQGS 479
DB 3 SAGGQGS 9

RESULT 51
ABP83398
ID ABP83398 standard; peptide; 18 AA.
XX AC ABP83398;
XX DT 04-MAR-2003 (first entry)
XX DE G protein-coupled receptor (GPCR) antigenic peptide SEQ ID NO:2071.
XX KW G protein-coupled receptor; GPCR; antigenic peptide; gene therapy;
KW G protein-coupled receptor modulator; antibody; immune-related disease;
KW growth-related disease; cell regeneration-related disease; AIDS; cancer;
KW immunological-related cell proliferative disease; autoimmune disease;
KW Alzheimer's disease; atherosclerosis; infection; osteoarthritis; allergy;
KW osteoporosis; cardiomyopathy; inflammation; Crohn's disease; diabetes;
KW graft-versus host disease; Parkinson's disease; multiple sclerosis; pain;
KW psoriasis; anxiety; depression; schizophrenia; dementia; memory loss;
KW mental retardation; epilepsy; asthma; tuberculosis; obesity; nausea;
KW hypertension; hypotension; renal disorder; rheumatoid arthritis; trauma;
XX KW ulcer.
XX OS Homo sapiens.
XX PN WO200261087-A2.
XX PD 08-AUG-2002.
XX PF 19-DEC-2001; 2001WO-US050107.
XX PR 19-DEC-2000; 2000US-0257144P.
XX PA (LIFE-) LIFESPAN BIOSCIENCES INC.
XX PI Burner GC, Roush CL, Brown JP;
XX WPI; 2003-046718/04.
XX
XX New isolated antigenic peptides e.g., for G protein-coupled receptors
XX (GPCR), useful for diagnosing and designing drugs for treating conditions
XX in which GPCRs are involved, e.g. AIDS, Alzheimer's disease, cancer or
XX autoimmune diseases.
XX
XX Claim 1; Fig 2; 523pp; English.
XX
XX The present invention describes antigenic peptides (I) comprising: (a)
XX any one of 1601 sequences (see ABP82019 to ABP82619) of 12-24 amino
XX acids. Also described: (1) an assay for the detection of a particular G
XX protein-coupled receptor (GPCR) or a candidate polypeptide in a sample;
XX and (2) an isolated antibody having high specificity and high affinity or
XX avidity for a particular GPCR. (I) can be used as GPCR modulators and in
XX gene therapy. The antigenic peptides for GPCRs are useful in detecting an
XX antibody against a particular GPCR, and in the production of specific
XX antibodies. The peptides and antibodies are also useful for detecting the

CC presence or absence of corresponding GPCRs. The antigenic peptides for
CC GPCRs and antibodies are useful for diagnosing and designing drugs for
CC treating immune-related diseases, growth-related diseases, cell
CC regeneration-related disease, immunological-related cell proliferative
CC diseases, or autoimmune diseases, e.g. AIDS, Alzheimer's disease,
CC atherosclerosis, bacterial, fungal, protozoan or viral infections,
CC osteoarthritis, osteoporosis, cancer, cardiomyopathy, chronic and acute
CC inflammation, allergies, Crohn's disease, diabetes, graft versus host
CC disease, Parkinson's disease, multiple sclerosis, pain, psoriasis,
CC anxiety, depression, schizophrenia, dementia, mental retardation, memory
CC loss, epilepsy, asthma, tuberculosis, obesity, nausea, hypertension,
CC hypotension, renal disorders, rheumatoid arthritis, trauma, ulcers, or
CC any other disorder in which GPCRs are involved. The antibodies may be
CC used in immunoassays and immunodiagnosis. ABZ42523 to ABZ42869 encode
CC GPCR proteins given in ABP81675 to ABP82018, which are used in the
CC exemplification of the present invention
XX
SQ Sequence 18 AA;

Query Match 1.3%; Score 7; DB 6; Length 18;
Best Local Similarity 100.0%; Pred. No. 75;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 57 EELSRQL 63
DB 6 EELSRQL 12
|||||

RESULT 52
AAW03444
ID AAW03444 standard; peptide; 25 AA.
XX
AC AAW03444;
XX
DT 14-FEB-1997 (first entry)
XX
DE HDV antigenic peptide, delta25-49.
XX
XX Hepatitis delta virus; HDV; antigen; delta12-60(Y); alpha-helix;
XX coiled-coil dimerisation domain; epitope; multimer.
XX Synthetic.
OS
XX WO9620953-A2.
FN
XX 11-JUL-1996.
PD
XX 22-DEC-1995; 95WO-US016854.
PF
XX 30-DEC-1994; 94US-00366479.
PR
XX (UTNC-) UNIV NORTH CAROLINA.
PA
XX Lemon SM, Erickson BW, Wang JG, Rozzelle J;
PI WPI; 1996-333940/33.
DR
XX Synthetic peptide with Delta hepatitis virus antigenic activity - useful
XX to produce and detect viral antibodies in a mammal and for immunisation.
PT
XX Example 10; Page 29; 50pp; English.
PS
XX The sequences given in AAW03442-47 represent peptides derived from the
XX hepatitis delta virus (HDV) antigen (HDVg) sequence. These sequences
XX represent synthetic peptides with hepatitis delta virus (HDV) antigenic
XX activity. They were used to raise murine monoclonal antibodies which may
XX be used to immunise animals against HDV, and to detect anti-HDV
XX antibodies. This sequence corresponds to segment B of HDAG
XX
SQ Sequence 25 AA;

Query Match 1.3%; Score 7; DB 2; Length 25;
Best Local Similarity 100.0%; Pred. No. 99;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 404 KIKKLE 410
DB 15 KIKKLE 21
|||||

RESULT 53
AAV64949
ID AAY64949 standard; protein; 27 AA.
XX
AC AAY64949;
XX
DT 01-FEB-2000 (first entry)
XX
DE Human 5' EST related polypeptide SEQ ID NO:1110.
XX
KW Human; 5' EST; expressed sequence tag; secreted protein; diagnosis;
KW gene therapy; chromosome mapping; upstream regulatory sequence; forensic;
KW location; development; protein synthesis; stability; regulation;
KW identification.
XX Homo sapiens.
OS
XX WO953051-A2.
FN
XX 21-OCT-1999.
PD
XX 09-APR-1999; 99WO-IB000712.
PF
XX 09-APR-1998; 98US-00057719.
PR
XX 28-APR-1998; 98US-00069047.
XX
PA (GEST) GENSET.
XX
PI Dumas Milne Edwards J, Duclert A, Giordano J;
DR WPI; 2000-038446/03.
DR N-PSDB; AAZ42563.
XX
PT Novel secreted protein 5' expressed sequence tag sequences used in
PT diagnostic, forensic, gene therapy, and chromosome mapping procedures.
XX
PS Claim 3; Page 689; 837pp; English.
XX
CC AAZ42265 to AAZ43075 represent novel 5' expressed sequence tag (EST)
CC sequences, corresponding to human secreted proteins. AAY64651 to AAY65438
CC represent the EST-related proteins corresponding to AAZ42265 to AAZ43052.
CC The 5' ESTs can be used for producing secreted human gene products. They
CC can be used to identify and isolate 5' untranslated regions (UTRs) and
CC upstream regulatory regions which control the location, development
CC stage, rate, and quantity of protein synthesis, as well as stability of
CC mRNA. The ESTs are also useful as probes for chromosome mapping, and to
CC obtain full length cDNA clones. The ESTs can also be used in forensic
CC procedures to identify individuals, or in diagnostic procedures to
CC identify individuals having genetic diseases resulting from abnormal gene
CC expression. The products may also be used in gene therapy protocols. The
CC nucleic acids encoding signal peptides can be used for directing
CC extracellular secretion of a polypeptide or the insertion of a
CC polypeptide into a membrane, or importing a polypeptide into a cell. The
CC proteins encoded by the EST sequences may be useful in treating a variety
CC of human conditions. Secreted proteins have therapeutic value, and the
CC identification of new secreted proteins is valuable. AAZ42249 to AAZ42264
CC and AAY64644 to AAY64650 represent sequences used in the exemplification
CC of the present invention
XX
SQ Sequence 27 AA;

Query Match 1.3%; Score 7; DB 3; Length 27;
Best Local Similarity 100.0%; Pred. No. 1.1e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 493 GQAQPS 499

DB |||||
4 GAQAPSS 10

RESULT 54
AAW03445
ID AAW03445 standard; peptide; 32 AA.
XX AC AAW03445;
XX DT 14-FEB-1997 (first entry)
XX DE HDV antigenic peptide, delta18-49.
XX KW Hepatitis delta virus; HDV; antigen; delta12-60(Y); alpha-helix;
XX KW coiled-coil dimerisation domain; epitope; multimer.
XX OS Synthetic.

XX Key Location/Qualifiers
FH Peptide 8..32
FT /label= Segment B
FT /note= "Corresponds to HDAG residues 25-49"
XX PN W09620953-A2.
XX PD 11-JUL-1996.
XX PF 22-DEC-1995; 95WO-US016854.
XX PR 30-DEC-1994; 94US-00366479.
XX PA (YINC-) UNIV NORTH CAROLINA.
XX PI Lemon SM, Erickson BW, Wang JG, Rozzelle J;
XX PI WPI; 1996-333940/33.
XX PT Synthetic peptide with Delta hepatitis virus antigenic activity - useful
XX PT to produce and detect viral antibodies in a mammal and for immunisation.
XX PS Example 10; Page 29; 50pp; English.

XX The sequences given in AAW03442-47 represent peptides derived from the
XX hepatitis delta virus (HDV) antigen (HDAG) sequence. These sequences
XX represent synthetic peptides with hepatitis delta virus (HDV) antigenic
XX activity. They were used to raise murine monoclonal antibodies which may
XX be used to immunise animals against HDV, and to detect anti-HDV
XX antibodies. This sequence corresponds to most of segment A and all of
XX segment B of HDAG
XX Sequence 32 AA;

Query Match 1.3%; Score 7; DB 2; Length 32;
Best Local Similarity 100.0%; Pred. No. 1.2e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 404 KXIKKLE 410
DB 22 KXIKKLE 28
|||

RESULT 55
ABP97458
ID ABP97458 standard; peptide; 34 AA.
XX AC ABP97458;
XX DT 04-AUG-2003 (first entry)
XX DE HIV N36-binding peptide SC34-b.
XX KW HIV; human immunodeficiency virus; N36-binding peptide; drug composite;

KW reverse transcriptase inhibitor; HIV protease inhibitor;
KW chemokine receptor antagonist; drug targeting; AIDS;
KW acquired immunodeficiency syndrome;
KW highly active anti-retroviral therapy; HAAR; anti-HIV.
XX Synthetic.

XX Key Location/Qualifiers
FH Modified-site 1
FT /note= "N-terminal acetyl"
FT Modified-site 34
FT /note= "C-terminal amide"
XX PN W02003029284-A1.
XX PD 10-APR-2003.
XX PF 27-SEP-2002; 2002WO-JP010119.
XX PR 27-SEP-2001; 2001JP-00297963.
XX PA (FUJI) FUJII N.
XX PI Fujii N, Otake A, Matsuoka M;
XX PI WPI; 2003-333562/31.
XX DR Anti-HIV agents containing polypeptide with high affinity toward N36 of
XX PT human immunodeficiency virus, or/and its composites, in drug compositions
XX PT for preventing or treating AIDS or onset of AIDS.
XX PS Example 1 #2; Page 11; 44pp; Japanese.

XX The invention relates to peptides with the ability to bind to N36 of
XX human immunodeficiency virus (HIV). The peptides have high affinity for
XX HIV N36, and have potent anti-HIV activity and high water solubility. The
XX peptides may optionally be linked to a reverse transcriptase inhibitor,
XX HIV protease inhibitor or chemokine receptor antagonist to form a
XX composite, and may also be used to target drugs to HIV-infected cells.
XX Peptides and/or composites of the invention may be used to prevent or
XX treat the onset of AIDS in HIV positive individuals or AIDS patients,
XX either alone or in combination with AIDS highly active anti-retroviral
XX therapies (HAAR). Sequences ABP97457-ABP97461 represent N36-binding
XX peptides of the invention
XX Sequence 34 AA;

Query Match 1.3%; Score 7; DB 6; Length 34;
Best Local Similarity 100.0%; Pred. No. 1.3e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 403 TKKIKKL 409
DB 12 TKKIKKL 18
|||

RESULT 56
ABP97457
ID ABP97457 standard; peptide; 34 AA.
XX AC ABP97457;
XX DT 04-AUG-2003 (first entry)
XX DE HIV N36-binding peptide SC34-a.
XX KW HIV; human immunodeficiency virus; N36-binding peptide; drug composite;
XX KW reverse transcriptase inhibitor; HIV protease inhibitor;
XX KW chemokine receptor antagonist; drug targeting; AIDS;
XX KW acquired immunodeficiency syndrome;
XX KW highly active anti-retroviral therapy; HAAR; anti-HIV.
XX OS Synthetic.

```
XX FH Key Location/Qualifiers
FT Modified-site 1
FT /note= "The N-terminus is either acetylated, or is linked
FT via a CO-(CH2)3-CO-NH-(CH2)5-CO- linker to AZT"
FT Modified-site 2
FT /label= Nle
FT Modified-site 34
FT /note= "C-terminal amide"
XX WO2003029284-A1.
XX 10-APR-2003.
XX 27-SEP-2002; 2002WO-JP010119.
XX 27-SEP-2001; 2001JP-00297963.
XX (FUJI/) FUJII N.
XX Fujii N, Otake A, Matsuoka M;
XX WPI; 2003-333562/31.
XX Anti-HIV agents containing polypeptide with high affinity toward N36 of
PT human immunodeficiency virus, or/and its composites, in drug compositions
PT for preventing or treating AIDS or onset of AIDS.
XX Example 1 #1; Page 11; 44pp; Japanese.
XX The invention relates to peptides with the ability to bind to N36 of
CC human immunodeficiency virus (HIV). The peptides have high affinity for
CC HIV N36, and have potent anti-HIV activity and high water solubility. The
CC peptides may optionally be linked to a reverse transcriptase inhibitor,
CC HIV protease inhibitor or chemokine receptor antagonist to form a
CC composite, and may also be used to target drugs to HIV-infected cells.
CC peptides and/or composites of the invention may be used to prevent or
CC treat the onset of AIDS in HIV positive individuals or AIDS patients,
CC either alone or in combination with AIDS highly active anti-retroviral
CC therapies (HAART). Sequences ABP97457-ABP97461 represent N36-binding
CC peptides of the invention
XX SQ Sequence 34 AA;
Query Match 1.3%; Score 7; DB 6; Length 34;
Best Local Similarity 100.0%; Pred. No. 1.3e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 403 TKXIKKL 409
Db 12 TKXIKKL 18
RESULT 57
AAW03447
ID AAW03447 standard; peptide; 35 AA.
AC AAW03447;
XX 14-FEB-1997 (first entry)
XX HDV antigenic peptide, delta(C)28-60(Y).
XX Hepatitis delta virus; HDV; antigen; delta12-60(Y); alpha-helix;
XX coiled-coil dimerisation domain; epitope; multimer.
XX Synthetic.
XX Key Location/Qualifiers
FH Peptide 1..23
FT /label= Segment B
FT /note= "Corresponds to HDag residues Cys-28-49"
FT Peptide 24..34
```

```
FT FT /label= Segment_C
FT /note= "Corresponds to HDag residues 50-60"
XX WO9620953-A2.
XX 11-JUL-1996.
XX 22-DEC-1995; 95WO-US016854.
XX 30-DEC-1994; 94US-00366479.
XX (UYNC-) UNIV NORTH CAROLINA.
XX Lemon SM, Erickson BW, Wang JG, Rozzelle J;
XX WPI; 1996-333940/33.
XX Synthetic peptide with Delta hepatitis virus antigenic activity - useful
PT to produce and detect viral antibodies in a mammal and for immunisation.
XX Example 10; Page 29; 50pp; English.
XX The sequences given in AAW03442-47 represent peptides derived from the
CC hepatitis delta virus (HDV) antigen (HDag) sequence. These sequences
CC represent synthetic peptides with hepatitis delta virus (HDV) antigenic
CC activity. They were used to raise murine monoclonal antibodies which may
CC be used to immunise animals against HDV, and to detect anti-HDV
CC antibodies. This sequence corresponds to most of segment B and all of
CC segment C of HDag
XX SQ Sequence 35 AA;
Query Match 1.3%; Score 7; DB 2; Length 35;
Best Local Similarity 100.0%; Pred. No. 1.3e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 404 KXIKKLE 410
Db 13 KXIKKLE 19
RESULT 58
AAW03446
ID AAW03446 standard; peptide; 35 AA.
XX AAW03446;
XX 14-FEB-1997 (first entry)
XX HDV antigenic peptide, delta15-49.
XX Hepatitis delta virus; HDV; antigen; delta12-60(Y); alpha-helix;
XX coiled-coil dimerisation domain; epitope; multimer.
XX Synthetic.
XX Key Location/Qualifiers
FH Peptide 11..35
FT /label= Segment_B
FT /note= "Corresponds to HDag residues 25-49"
FT WO9620953-A2.
XX 11-JUL-1996.
XX 22-DEC-1995; 95WO-US016854.
XX 30-DEC-1994; 94US-00366479.
XX (UYNC-) UNIV NORTH CAROLINA.
XX Lemon SM, Erickson BW, Wang JG, Rozzelle J;
```

DR WPI; 1996-333940/33.
XX Synthetic peptide with Delta hepatitis virus antigenic activity - useful
PT to produce and detect viral antibodies in a mammal and for immunisation.
XX
PS Example 10; Page 29; 50pp; English.
XX
CC The sequences given in AA03442-47 represent peptides derived from the
CC hepatitis delta virus (HDV) antigen (HDAG) sequence. These sequences
CC represent synthetic peptides with hepatitis delta virus (HDV) antigenic
CC activity. They were used to raise murine monoclonal antibodies which may
CC be used to immunise animals against HDV, and to detect anti-HDV
CC antibodies. This sequence corresponds to most of segment A and all of
CC segment B of HDAG
XX
SQ Sequence 35 AA;
Query Match 1.3%; Score 7; DB 2; Length 35;
Best Local Similarity 100.0%; Pred. No. 1.3e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 404 KRIKLE 410
Db 25 KRIKLE 31
RESULT 59
ABG23756
ID ABG23756 standard; protein; 36 AA.
XX
AC ABG23756;
XX
DT 18-FEB-2002 (first entry)
XX
DE Novel human diagnostic protein #23747.
XX
KW Human; chromosome mapping; gene mapping; gene therapy; forensic;
KW food supplement; medical imaging; diagnostic; genetic disorder.
XX
OS Homo sapiens.
XX
PN WO200175067-A2.
XX
PD 11-OCT-2001.
XX
PF 30-MAR-2001; 2001WO-US008631.
XX
PR 31-MAR-2000; 2000US-00540217.
XX
PR 23-AUG-2000; 2000US-00649167.
XX
PA (HYSE-) HYSEQ INC.
XX
PI Drmanac RT, Liu C, Tang YT;
XX
DR WPI; 2001-639362/73.
DR N-PSDB; AAS87943.
XX
PT New isolated polynucleotide and encoded polypeptides, useful in
PT diagnostics, forensics, gene mapping, identification of mutations
PT responsible for genetic disorders or other traits and to assess
PT biodiversity.
XX
PS Claim 20; SEQ ID NO 54125; 103pp; English.
XX
CC The invention relates to isolated polynucleotide (I) and polypeptide (II)
CC sequences. (I) is useful as hybridisation probes, polymerase chain
CC reaction (PCR) primers, oligomers, and for chromosome and gene mapping,
CC and in recombinant production of (II). The polynucleotides are also used
CC in diagnostics as expressed sequence tags for identifying expressed
CC genes. (I) is useful in gene therapy techniques to restore normal
CC activity of (I) or to treat disease states involving (II). (II) is
CC useful for generating antibodies against it, detecting or quantitating a
CC polypeptide in tissue, as molecular weight markers and as a food

CC supplement. (II) and its binding partners are useful for treating disorders
CC of sites expressing (II). (I) and (II) are useful for treating disorders
CC involving aberrant protein expression or biological activity. The
CC polypeptide and polynucleotide sequences have applications in
CC diagnostics, forensics, gene mapping, identification of mutations
CC and to produce other types of data and products dependent on DNA and
CC amino acid sequences. ABG00010-ABG0377 represent novel human diagnostic
CC amino acid sequences of the invention. Note: The sequence data for this
CC patent did not appear in the printed specification, but was obtained in
CC electronic format directly from WIPO at
CC ftp.wipo.int/pub/published_pct_sequences
XX
SQ Sequence 36 AA;
Query Match 1.3%; Score 7; DB 4; Length 36;
Best Local Similarity 100.0%; Pred. No. 1.3e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 477 QGSLTDS 483
Db 19 QGSLTDS 25
RESULT 60
ABG27966
ID ABG27966 standard; protein; 36 AA.
XX
AC ABG27966;
XX
DT 18-FEB-2002 (first entry)
XX
DE Novel human diagnostic protein #27957.
XX
KW Human; chromosome mapping; gene mapping; gene therapy; forensic;
KW food supplement; medical imaging; diagnostic; genetic disorder.
XX
OS Homo sapiens.
XX
PN WO200175067-A2.
XX
PD 11-OCT-2001.
XX
PF 30-MAR-2001; 2001WO-US008631.
XX
PR 31-MAR-2000; 2000US-00540217.
XX
PR 23-AUG-2000; 2000US-00649167.
XX
PA (HYSE-) HYSEQ INC.
XX
PI Drmanac RT, Liu C, Tang YT;
XX
DR WPI; 2001-639362/73.
DR N-PSDB; AAS92153.
XX
PT New isolated polynucleotide and encoded polypeptides, useful in
PT diagnostics, forensics, gene mapping, identification of mutations
PT responsible for genetic disorders or other traits and to assess
PT biodiversity.
XX
PS Claim 20; SEQ ID NO 58325; 103pp; English.
XX
CC The invention relates to isolated polynucleotide (I) and polypeptide (II)
CC sequences. (I) is useful as hybridisation probes, polymerase chain
CC reaction (PCR) primers, oligomers, and for chromosome and gene mapping,
CC and in recombinant production of (II). The polynucleotides are also used
CC in diagnostics as expressed sequence tags for identifying expressed
CC genes. (I) is useful in gene therapy techniques to restore normal
CC activity of (I) or to treat disease states involving (II). (II) is
CC useful for generating antibodies against it, detecting or quantitating a
CC polypeptide in tissue, as molecular weight markers and as a food
CC supplement. (II) and its binding partners are useful in medical imaging
CC of sites expressing (II). (I) and (II) are useful for treating disorders


```
ABG27959
ID  ABG27959 standard; protein; 37 AA.
XX  AC
XX  ABG27959;
XX  DT
XX  18-FEB-2002 (first entry)
XX  DE
XX  Novel human diagnostic protein #27950.
XX  Human; chromosome mapping; gene mapping; gene therapy; forensic;
XX  food supplement; medical imaging; diagnostic; genetic disorder.
XX  KW
XX  OS
XX  Homo sapiens.
XX  FN
XX  WO200175067-A2.
XX  PD
XX  11-OCT-2001.
XX  PF
XX  30-MAR-2001; 2001WO-US008631.
XX  PR
XX  31-MAR-2000; 2000US-00540217.
XX  PR
XX  23-AUG-2000; 2000US-00649167.
XX  PA
XX  (HYSE-) HYSEQ INC.
XX  PI
XX  Drmanac RT, Liu C, Tang YT;
XX  WPI; 2001-639362/73.
XX  DR
XX  N-PSDB; AAS92146.
XX  New isolated polynucleotide and encoded polypeptides, useful in
XX  diagnostics, forensics, gene mapping, identification of mutations
XX  responsible for genetic disorders or other traits and to assess
XX  biodiversity.
XX  Claim 20; SEQ ID NO 58318; 103pp; English.
XX  The invention relates to isolated polynucleotide (I) and polypeptide (II)
XX  sequences. (I) is useful as hybridisation probes, polymerase chain
XX  reaction (PCR) primers, oligomers, and for chromosome and gene mapping,
XX  and in recombinant production of (II). The polynucleotides are also used
XX  in diagnostics as expressed sequence tags for identifying expressed
XX  genes. (I) is useful in gene therapy techniques to restore normal
XX  activity of (II) or to treat disease states involving (II). (II) is
XX  useful for generating antibodies against it, detecting or quantitating a
XX  polypeptide in tissue, as molecular weight markers and as a food
XX  supplement. (II) and its binding partners are useful in medical imaging
XX  of sites expressing (II). (I) and (II) are useful for treating disorders
XX  involving aberrant protein expression or biological activity. The
XX  polypeptide and polynucleotide sequences have applications in
XX  diagnostics, forensics, gene mapping, identification of mutations
XX  responsible for genetic disorders or other traits to assess biodiversity
XX  and to produce other types of data and products dependent on DNA and
XX  amino acid sequences. ABG0010-ABG30377 represent novel human diagnostic
XX  amino acid sequences of the invention. Note: The sequence data for this
XX  patent did not appear in the printed specification, but was obtained in
XX  electronic format directly from WIPO at
XX  ftp.wipo.int/pub/published_pct_sequences
XX  SQ
XX  Sequence 37 AA;
XX  Query Match 1.3%; Score 7; DB 4; Length 37;
XX  Best Local Similarity 100.0%; Pred. No. 1.4e+02;
XX  Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
XX  QY 477 QGSLTDS 483
XX  Db 20 QGSLTDS 26
XX  RESULT 64
XX  AAW03442
XX  ID AAW03442 standard; peptide; 38 AA.
```

```
XX  AAW03442;
XX  AC
XX  14-FEB-1997 (first entry)
XX  DT
XX  HDV antigenic peptide, delta12-49.
XX  DE
XX  Hepatitis delta virus; HDV; antigen; delta12-60(Y); alpha-helix;
XX  KW coiled-coil dimerisation domain; epitope; multimer.
XX  OS
XX  Synthetic.
XX  FN
XX  WO9620953-A2.
XX  PD
XX  11-JUL-1996.
XX  PF
XX  22-DEC-1995; 95WO-US016854.
XX  PR
XX  30-DEC-1994; 94US-00366479.
XX  PA
XX  (UTNC-) UNIV NORTH CAROLINA.
XX  PI
XX  Lemon SM, Erickson BW, Wang JG, Rozzelle J;
XX  WPI; 1996-333940/33.
XX  DR
XX  Synthetic peptide with Delta hepatitis virus antigenic activity - useful
XX  to produce and detect viral antibodies in a mammal and for immunisation.
XX  Example 10; Page 29; 50pp; English.
XX  The sequences given in AAW03442-47 represent peptides derived from the
XX  hepatitis delta virus (HDV) antigen (HDag) sequence. These sequences
XX  represent synthetic peptides with hepatitis delta virus (HDV) antigenic
XX  activity. They were used to raise murine monoclonal antibodies which may
XX  be used to immunise animals against HDV, and to detect anti-HDV
XX  antibodies
XX  SQ
XX  Sequence 38 AA;
XX  Query Match 1.3%; Score 7; DB 2; Length 38;
XX  Best Local Similarity 100.0%; Pred. No. 1.4e+02;
XX  Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
XX  QY 404 KKIKKLE 410
XX  Db 28 KKIKKLE 34
XX  RESULT 65
XX  AAW16323
XX  ID AAW16323 standard; protein; 38 AA.
XX  AC
XX  AAW16323;
XX  DT
XX  12-OCT-2001 (first entry)
XX  DE
XX  Peptide #2757 encoded by probe for measuring cervical gene expression.
XX  KW Probe; human; microarray; gene expression; cervical epithelial cell;
XX  cervical cancer.
XX  OS
XX  Homo sapiens.
XX  FN
XX  WO200157278-A2.
XX  PD
XX  09-AUG-2001.
XX  PF
XX  30-JAN-2001; 2001WO-US000670.
XX  PR
XX  04-FEB-2000; 2000US-0180312P.
XX  PR
XX  26-MAY-2000; 2000US-0207456P.
XX  PR
XX  30-JUN-2000; 2000US-00608408.
```

```
PT 03-AUG-2000; 2000US-00632366.
XX 21-SEP-2000; 2000US-0234687P.
PS 27-SEP-2000; 2000US-0236359P.
XX 04-OCT-2000; 2000GB-00024263.
XX (MOLE-) MOLECULAR DYNAMICS INC.
XX Penn SG, Hanzel DK, Chen W, Rank DR;
XX WPI; 2001-488901/53.
XX Human genome-derived single exon nucleic acid probes useful for analyzing
PT gene expression in human cervical epithelial cells.
XX Claim 27; SEQ ID NO 21149; 487pp; English.
XX The present invention relates to human single exon nucleic acid probes
CC (SENP; see AA110069-AA128459). The present sequence is a peptide encoded
CC by one such probe. The SENPs are derived from human HeLa cells. The SENPs
CC can be used to produce a single exon microarray, which can be used for
CC measuring human gene expression in a sample derived from human cervical
CC epithelial cells. By measuring gene expression, the probes are therefore
CC useful in grading and/or staging of diseases of the cervix, notably
CC cervical cancer. Note: The sequence data for this patent did not form
CC part of the printed specification, but was obtained in electronic format
CC directly from WIPO at ftp.wipo.int/pub/published_pct_sequences
XX Sequence 38 AA;
XX
XX Query Match 1.3%; Score 7; DB 4; Length 38;
XX Best Local Similarity 100.0%; Pred. No. 1.4e+02;
XX Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
XX
XX Qy 281 KLRQENM 287
XX Db 23 KLRQENM 29
XX |||||
XX 23 KLRQENM 29
XX
XX RESULT 66
XX ABB35315
XX ID ABB35315 standard; peptide; 38 AA.
XX AC ABB35315;
XX DT 04-FEB-2002 (first entry)
XX DE Peptide #2821 encoded by human foetal liver single exon probe.
XX KW Human; foetal liver; gene expression; single exon nucleic acid probe.
XX OS Homo sapiens.
XX PN WO200157277-A2.
XX PD 09-AUG-2001.
XX PF 30-JAN-2001; 2001WO-US000669.
XX PR 04-FEB-2000; 2000US-0180312P.
XX PR 26-MAY-2000; 2000US-0207456P.
XX PR 30-JUN-2000; 2000US-00608408.
XX PR 03-AUG-2000; 2000US-00632366.
XX PR 21-SEP-2000; 2000US-0234687P.
XX PR 27-SEP-2000; 2000US-0236359P.
XX PR 04-OCT-2000; 2000GB-00024263.
XX PA (MOLE-) MOLECULAR DYNAMICS INC.
XX PI Penn SG, Hanzel DK, Chen W, Rank DR;
XX WPI; 2001-483447/52.
XX Human genome-derived single exon nucleic acid probes useful for analyzing
PT
XX 03-AUG-2000; 2000US-00632366.
XX 21-SEP-2000; 2000US-0234687P.
XX 27-SEP-2000; 2000US-0236359P.
XX 04-OCT-2000; 2000GB-00024263.
XX (MOLE-) MOLECULAR DYNAMICS INC.
XX Penn SG, Hanzel DK, Chen W, Rank DR;
XX WPI; 2001-488901/53.
XX Human genome-derived single exon nucleic acid probes useful for measuring
PT gene expression in human foetal liver.
XX Claim 27; SEQ ID NO 27950; 639pp + Sequence Listing; English.
XX The invention relates to a single exon nucleic acid probe for measuring
CC human gene expression in a sample derived from human foetal liver. The
CC single exon nucleic acid probes may be used for predicting, measuring and
CC displaying gene expression in samples derived from human foetal liver. The
CC present sequence is a peptide encoded by a single exon nucleic acid probe
CC of the invention. Note: The sequence data for this patent did not form
CC part of the printed specification, but was obtained in electronic format
CC directly from WIPO at ftp.wipo.int/pub/published_pct_sequences
XX Sequence 38 AA;
XX
XX Query Match 1.3%; Score 7; DB 4; Length 38;
XX Best Local Similarity 100.0%; Pred. No. 1.4e+02;
XX Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
XX
XX Qy 281 KLRQENM 287
XX Db 23 KLRQENM 29
XX |||||
XX 23 KLRQENM 29
XX
XX RESULT 67
XX AAM28818
XX ID AAM28818 standard; protein; 38 AA.
XX AC AAM28818;
XX DT 17-OCT-2001 (first entry)
XX DE Peptide #2855 encoded by probe for measuring placental gene expression.
XX KW Probe; microarray; human; placenta; antenatal diagnosis;
XX genetic disorder.
XX OS Homo sapiens.
XX PN WO200157272-A2.
XX PD 09-AUG-2001.
XX PF 30-JAN-2001; 2001WO-US000663.
XX PR 04-FEB-2000; 2000US-0180312P.
XX PR 26-MAY-2000; 2000US-0207456P.
XX PR 30-JUN-2000; 2000US-00608408.
XX PR 03-AUG-2000; 2000US-00632366.
XX PR 21-SEP-2000; 2000US-0234687P.
XX PR 27-SEP-2000; 2000US-0236359P.
XX PR 04-OCT-2000; 2000GB-00024263.
XX PA (MOLE-) MOLECULAR DYNAMICS INC.
XX PI Penn SG, Hanzel DK, Chen W, Rank DR;
XX WPI; 2001-488907/53.
XX Human genome-derived single exon nucleic acid probes useful for analyzing
PT gene expression in human placenta.
XX Claim 27; SEQ ID NO 23087; 654pp; English.
XX The present invention relates to single exon nucleic acid probes (SENP;
CC see AA131315-AA157546). The present sequence is a peptide encoded by one
CC such probe. The probes are useful for producing a microarray for
CC predicting, measuring and displaying gene expression in samples derived
CC from human placenta. The probes are useful for antenatal diagnosis of
CC human genetic disorders
XX Sequence 38 AA;
XX
```

Query Match 1.3%; Score 7; DB 4; Length 38;
Best Local Similarity 100.0%; Pred. No. 1.4e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 281 KLRQENM 287
|||
Db 23 KLRQENM 29

RESULT 68
ABB30145
ID ABB30145 standard; peptide; 38 AA.

XX AC
XX ABB30145;

XX 01-FEB-2002 (first entry)

DE Peptide #2796 encoded by breast cell single exon nucleic acid probe.

XX Human; microarray; single exon probe; gene expression; breast; disease;
cancer.

XX OS
XX Homo sapiens.

XX PN
XX WO200157271-A2.

XX 09-AUG-2001.

XX 30-JAN-2001; 2001WO-US000662.

XX 04-FEB-2000; 2000US-0180312P.

XX 26-MAY-2000; 2000US-0207456P.

XX 30-JUN-2000; 2000US-00608408.

XX 03-AUG-2000; 2000US-00632366.

XX 21-SEP-2000; 2000US-0234687P.

XX 27-SEP-2000; 2000US-0236359P.

XX 04-OCT-2000; 2000GB-00024263.

XX (MOLE-) MOLECULAR DYNAMICS INC.

XX PI
XX Penn SG, Hanzel DK, Chen W, Rank DR;

XX WPI; 2001-496933/54.

XX New spatially-addressable set of single exon nucleic acid probes, useful
PT for measuring gene expression in sample derived from human breast,
PT comprises number of single exon nucleic acid probes.

XX Claim 27; SEQ ID NO 13113; 327pp + Sequence Listing; English.

XX CC The invention relates to a spatially-addressable set of single exon
CC nucleic acid probes for measuring gene expression in a sample derived
CC from human breast and BT 474 cells. The method involves contacting the
CC probes with a collection of detectably labelled nucleic acids derived
CC from mRNA of human breast, and then measuring the label bound to each
CC probe of the microarray. The probes are useful for verifying the
CC expression of regions of genomic DNA predicted to encode proteins. They
CC are useful for gene discovery, and for determining predisposition and/or
CC prognosing breast disease. Gene expression analysis is useful for
CC assessing the toxicity of chemical agents on cells. The microarray of
CC this invention presents a far greater diversity of probes for measuring
CC gene expression, with far less bias than expressed sequence tag
CC microarrays. The method is suitable for rapid production of functional
CC information from genomic sequence. The present sequence is a peptide
CC encoded by a single exon nucleic acid probe of the invention. Note: The
CC sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences

XX SQ Sequence 38 AA;

Query Match 1.3%; Score 7; DB 4; Length 38;
Best Local Similarity 100.0%; Pred. No. 1.4e+02;

Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 281 KLRQENM 287
|||
Db 23 KLRQENM 29

RESULT 69

ABB20757
ID ABB20757 standard; protein; 38 AA.

XX AC
XX ABB20757;

XX 23-JAN-2002 (first entry)

DE Protein #2756 encoded by probe for measuring heart cell gene expression.

XX Human; gene expression; heart; microarray; vascular system;
XX cardiovascular disease; hypertension; cardiac arrhythmia;
XX congenital heart disease.

XX OS
XX Homo sapiens.

XX PN
XX WO200157274-A2.

XX 09-AUG-2001.

XX 30-JAN-2001; 2001WO-US000666.

XX 04-FEB-2000; 2000US-0180312P.

XX 26-MAY-2000; 2000US-0207456P.

XX 30-JUN-2000; 2000US-00608408.

XX 03-AUG-2000; 2000US-00632366.

XX 21-SEP-2000; 2000US-0234687P.

XX 27-SEP-2000; 2000US-0236359P.

XX 04-OCT-2000; 2000GB-00024263.

XX (MOLE-) MOLECULAR DYNAMICS INC.

XX PI
XX Penn SG, Hanzel DK, Chen W, Rank DR;

XX WPI; 2001-488999/53.

XX Single exon nucleic acid probes for analyzing gene expression in human
PT hearts.

XX Claim 15; SEQ ID NO 22527; 530pp; English.

XX CC The present invention relates to single exon nucleic acid probes for
CC measuring human gene expression in a sample derived from human heart (see
CC ABA21535-ABA41305). The present sequence is a protein encoded by one such
CC probe. The probes may be used for predicting, measuring and displaying
CC gene expression in samples derived from the human heart via microarrays.
CC By measuring gene expression, the probes are useful for predicting,
CC diagnosing, grading, staging, monitoring and prognosing diseases of the
CC human heart and vascular system e.g. cardiovascular disease.
CC hypertension, cardiac arrhythmias and congenital heart disease. Note: The
CC sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences

XX SQ Sequence 38 AA;

Query Match 1.3%; Score 7; DB 4; Length 38;
Best Local Similarity 100.0%; Pred. No. 1.4e+02;

Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 281 KLRQENM 287
|||
Db 23 KLRQENM 29

RESULT 70

XX Penn SG, Hanzel DK, Chen W, Rank DR;
XX WPI; 2001-48898/53.
XX Human genome-derived single exon nucleic acid probes useful for analyzing
XX gene expression in human adult liver.
XX Claim 27; SEQ ID NO 28836; 658bp; English.
XX The invention relates to a single exon nucleic acid probe (SENP) (I) for
XX measuring human gene expression in a sample derived from human adult
XX liver, comprising one of 13109 defined nucleotide sequences given in the
XX specification (or complements/ fragments). The probe hybridises at high
XX stringency to a nucleic acid molecule expressed in the human adult liver.
XX (I) may be used for predicting, measuring and displaying gene expression
XX in samples derived from human adult liver. The genes identified may be
XX involved in genetic liver diseases such as cirrhosis,
XX hyperlipoproteinaemia, hyperlipidaemia and hypercholesterolaemia which is
XX associated with coronary heart disease. ABG47348-ABG59930 represent human
XX liver single exon encoded peptides of the invention. Note: The sequence
XX information for this patent does not appear in the printed specification
XX but was obtained in electronic format directly from WIPO at
XX ftp.wipo.int/pub/published_pct_sequences
XX Sequence 38 AA;
XX
XX Query Match 1.3%; Score 7; DB 4; Length 38;
XX Best Local Similarity 100.0%; Pred. No. 1.4e+02;
XX Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
XX
XX QY 281 KLRQENM 287
XX DB 23 KLRQENM 29
XX
XX RESULT 73
XX AA04061
XX ID AA04061 standard; protein; 38 AA.
XX AC AA04061;
XX DT 09-OCT-2001 (first entry)
XX DE Peptide #2743 encoded by probe for measuring breast gene expression.
XX KW Probe; human; breast disease; breast cancer; development disorder;
XX KW inflammatory disease; proliferative breast disease; non-carcinoma tumour.
XX OS Homo sapiens.
XX PN WO200157270-A2.
XX PD 09-AUG-2001.
XX PF 29-JAN-2001; 2001WO-US000661.
XX PR 04-FEB-2000; 2000US-0180312P.
XX PR 26-MAY-2000; 2000US-0207456P.
XX PR 30-JUN-2000; 2000US-00608408.
XX PR 03-AUG-2000; 2000US-00632366.
XX PR 21-SEP-2000; 2000US-0234687P.
XX PR 27-SEP-2000; 2000US-0236359P.
XX PR 04-OCT-2000; 2000GB-00024263.
XX (MOLE-) MOLECULAR DYNAMICS INC.
XX Penn SG, Hanzel DK, Chen W, Rank DR;
XX WPI; 2001-476286/51.
XX Novel single exon nucleic acid probe used to measuring gene expression in
XX a human breast.

XX PS Claim 27; SEQ ID NO 12801; 322pp; English.
XX The present invention relates to novel single exon nucleic acid probes
XX (see AA00010-AA11067). The present sequence is a peptide encoded by one
XX such probe. The probes are useful for measuring human gene expression in
XX a human breast sample, where the probe hybridises at high stringency to a
XX nucleic acid expressed in the human breast. The probes are useful for
XX predicting, diagnosing, grading, staging, monitoring and prognosing
XX diseases of the human breast, particularly those diseases with polygenic
XX aetiology. The diseases include: breast cancer, disorders of development,
XX inflammatory diseases of the breast, fibrocystic changes, proliferative
XX breast disease and non-carcinoma tumours. Note: The sequence data for
XX this patent did not form part of the printed specification, but was
XX obtained in electronic format directly from WIPO at
XX ftp.wipo.int/pub/published_pct_sequences
XX Sequence 38 AA;
XX
XX Query Match 1.3%; Score 7; DB 4; Length 38;
XX Best Local Similarity 100.0%; Pred. No. 1.4e+02;
XX Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
XX
XX QY 281 KLRQENM 287
XX DB 23 KLRQENM 29
XX
XX RESULT 74
XX ABG38100
XX ID ABG38100 standard; peptide; 38 AA.
XX AC ABG38100;
XX DT 19-AUG-2002 (first entry)
XX DE Human peptide encoded by genome-derived single exon probe SEQ ID 27765.
XX KW Human; single exon probe; asthma; lung cancer; COPD; ILD;
XX KW chronic obstructive pulmonary disease; interstitial lung disease;
XX KW familial idiopathic pulmonary fibrosis; neurofibromatosis;
XX KW tuberous sclerosis; Gaucher's disease; Niemann-Pick disease;
XX KW Hermansky-Pudlak syndrome; sarcoidosis; pulmonary haemosiderosis;
XX KW pulmonary histiocytosis; lymphangioleiomyomatosis; Karagener syndrome;
XX KW pulmonary alveolar proteinosis; fibrocystic pulmonary dysplasia;
XX KW primary ciliary dyskinesia; pulmonary hypertension;
XX KW hyaline membrane disease.
XX OS Homo sapiens.
XX PN WO200186003-A2.
XX PD 15-NOV-2001.
XX PF 30-JAN-2001; 2001WO-US000665.
XX PR 04-FEB-2000; 2000US-0180312P.
XX PR 26-MAY-2000; 2000US-0207456P.
XX PR 30-JUN-2000; 2000US-00608408.
XX PR 03-AUG-2000; 2000US-00632366.
XX PR 21-SEP-2000; 2000US-0234687P.
XX PR 27-SEP-2000; 2000US-0236359P.
XX PR 04-OCT-2000; 2000GB-00024263.
XX (MOLE-) MOLECULAR DYNAMICS INC.
XX Penn SG, Hanzel DK, Chen W, Rank DR;
XX WPI; 2002-114183/15.
XX Spatially-addressable set of single exon nucleic acid probes, used to
XX measure gene expression in human lung samples.

PS Claim 27; SEQ ID NO 27765; 634pp; English.

XX The invention relates to a spatially-addressable set of single exon
 CC nucleic acid probes for measuring gene expression in a sample derived
 CC from human lung comprising single exon nucleic acid probes having one of
 CC 12614 nucleic acid sequences mentioned in the specification, or their
 CC complements or the 12387 open reading frames derived from the 12614
 CC probes. Also included are a microarray comprising the novel set of probes
 CC; the novel set of probes which hybridise at high stringency to a nucleic
 CC acid expressed in the human lung; measuring gene expression in a sample
 CC derived from human lung, comprising (a) contacting the array with a
 CC collection of detectably labeled nucleic acids derived from human lung
 CC mRNA; and (b) measuring the label detectably bound to each probe of the
 CC array; identifying exons in a eukaryotic genome, comprising (a)
 CC algorithmically predicting at least one exon from genomic sequences of
 CC the eukaryote; and (b) detecting specific hybridisation of detectably
 CC labeled nucleic acids from eukaryote lung mRNA, to a single exon probe,
 CC having a fragment identical to the predicted exon, the probe is included
 CC in the above mentioned microarray; assigning exons to a single gene,
 CC comprising (a) identifying exons from genomic sequence by the method
 CC above and (b) measuring the expression of each of the exons in several
 CC tissues and/or cell types using hybridisation to a single exon
 CC microarrays having a probe with the exon, where a common pattern of
 CC expression of the exons in the tissues and/or cell types indicates that
 CC the exons should be assigned to a single gene; a peptide comprising one
 CC of 12011 sequences, mentioned in the specification, or encoded by the
 CC probes/open reading frames (ORF). The probes are used for gene expression
 CC analysis, and for identifying exons in a gene, particularly using human
 CC lung derived mRNA and for the study of lung diseases such as asthma, lung
 CC cancer, chronic obstructive pulmonary disease (COPD), interstitial lung
 CC disease (ILD), familial idiopathic pulmonary fibrosis, neurofibromatosis,
 CC tuberous sclerosis, Gaucher's disease, Niemann-Pick disease, Hermansky-
 CC Pudlak syndrome, sarcoidosis, pulmonary haemosiderosis, pulmonary
 CC histiocytosis, lymphangioleiomyomatosis, pulmonary alveolar proteinosis,
 CC Karagenar syndrome, fibrocystic pulmonary dysplasia, primary ciliary
 CC dyskinesia, pulmonary hypertension and hyaline membrane disease. The
 CC present sequence is a peptide/protein encoded by a single exon probe of
 CC the invention. Note: The sequence data for this patent did not form part
 CC of the printed specification, but was obtained in electronic format
 CC directly from WIPO at ftp.wipo.int/pub/published_pct_sequences

XX
 SQ Sequence 33 AA;

Query Match 1.3%; Score 7; DB 5; Length 38;
 Best Local Similarity 100.0%; Pred. No. 1.4e+02;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 281 KLRQNM 287
 DB 23 KLRQNM 29
 |||||

RESULT 75
 ABG28175
 ID ABG28175 standard; protein; 42 AA.
 XX
 AC ABG28175;
 XX
 DT 18-FEB-2002 (first entry)
 XX
 DE Novel human diagnostic protein #28166.
 XX
 KW Human; chromosome mapping; gene mapping; gene therapy; forensic;
 KW food supplement; medical imaging; diagnostic; genetic disorder.
 XX
 OS Homo sapiens.
 XX
 PN WO200175067-A2.
 XX
 PD 11-OCT-2001.
 XX
 PF 30-MAR-2001; 2001WO-US008631.
 XX
 PR 31-MAR-2000; 2000US-00540217.
 PR 23-AUG-2000; 2000US-00649167.

PR 31-MAR-2000; 2000US-00540217.
 PR 23-AUG-2000; 2000US-00649167.
 XX (HYSE-) HYSEQ INC.
 XX Drmanac RT, Liu C, Tang YT;
 XX WPI: 2001-639362/73.
 DR N-PSDB; AAS92362.
 DR
 XX New isolated polynucleotide and encoded polypeptides, useful in
 PT diagnostics, forensics, gene mapping, identification of mutations
 PT responsible for genetic disorders or other traits and to assess
 PT biodiversity.
 XX Claim 20; SEQ ID NO 58534; 103pp; English.
 XX The invention relates to isolated polynucleotide (I) and polypeptide (II)
 CC sequences. (I) is useful as hybridisation probes, polymerase chain
 CC reaction (PCR) primers, oligomers, and for chromosome and gene mapping,
 CC and in recombinant production of (II). The polynucleotides are also used
 CC in diagnostics as expressed sequence tags for identifying expressed
 CC genes. (I) is useful in gene therapy techniques to restore normal
 CC activity of (II) or to treat disease states involving (II). (II) is
 CC useful for generating antibodies against it, detecting or quantitating a
 CC polypeptide in tissue, as molecular weight markers and as a food
 CC supplement. (II) and its binding partners are useful in medical imaging
 CC of sites expressing (II). (I) and (II) are useful for treating disorders
 CC involving aberrant protein expression or biological activity. The
 CC polypeptide and polynucleotide sequences have applications in
 CC diagnostics, forensics, gene mapping, identification of mutations
 CC responsible for genetic disorders or other traits to assess biodiversity
 CC and to produce other types of data and products dependent on DNA and
 CC amino acid sequences. ABG0010-ABG30377 represent novel human diagnostic
 CC amino acid sequences of the invention. Note: The sequence data for this
 CC patent did not appear in the printed specification, but was obtained in
 CC electronic format directly from WIPO at
 CC ftp.wipo.int/pub/published_pct_sequences

XX
 SQ Sequence 42 AA;

Query Match 1.3%; Score 7; DB 4; Length 42;
 Best Local Similarity 100.0%; Pred. No. 1.5e+02;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 477 QGSLTDS 483
 DB 25 QGSLTDS 31
 |||||

RESULT 76
 ABG23760
 ID ABG23760 standard; protein; 43 AA.
 XX
 AC ABG23760;
 XX
 DT 18-FEB-2002 (first entry)
 XX
 DE Novel human diagnostic protein #23751.
 XX
 KW Human; chromosome mapping; gene mapping; gene therapy; forensic;
 KW food supplement; medical imaging; diagnostic; genetic disorder.
 XX
 OS Homo sapiens.
 XX
 PN WO200175067-A2.
 XX
 PD 11-OCT-2001.
 XX
 PF 30-MAR-2001; 2001WO-US008631.
 XX
 PR 31-MAR-2000; 2000US-00540217.
 PR 23-AUG-2000; 2000US-00649167.

XX PA (HYSE-) HYSEQ INC.

XX PI Drmanac RT, Liu C, Tang YT;

XX DR WPI; 2001-639362/73.

XX DR N-PSDB; AAS87947.

XX DR New isolated polynucleotide and encoded polypeptides, useful in

XX PT diagnostics, forensics, gene mapping, identification of mutations

XX PT responsible for genetic disorders or other traits and to assess

XX PT biodiversity.

XX PS Claim 20; SEQ ID NO 54119; 103pp; English.

XX CC The invention relates to isolated polynucleotide (I) and polypeptide (II)

XX CC sequences. (I) is useful as hybridisation probes, polymerase chain

XX CC reaction (PCR) primers, oligomers, and for chromosome and gene mapping,

XX CC and in recombinant production of (II). The polynucleotides are also used

XX CC in diagnostics as expressed sequence tags for identifying expressed

XX CC genes. (I) is useful in gene therapy techniques to restore normal

XX CC activity of (II) or to treat disease states involving (II). (II) is

XX CC useful for generating antibodies against it, detecting or quantitating a

XX CC polypeptide in tissue, as molecular weight markers and as a food

XX CC supplement. (II) and its binding partners are useful in medical imaging

XX CC of sites expressing (II). (I) and (II) are useful for treating disorders

XX CC involving aberrant protein expression or biological activities. The

XX CC polypeptide and polynucleotide sequences have applications in

XX CC diagnostics, forensics, gene mapping, identification of mutations

XX CC responsible for genetic disorders or other traits to assess biodiversity

XX CC and to produce other types of data and products dependent on DNA and

XX CC amino acid sequences. ABG00010-ABG30377 represent novel human diagnostic

XX CC amino acid sequences of the invention. Note: The sequence data for this

XX CC patent did not appear in the printed specification, but was obtained in

XX CC electronic format directly from WIPO at

XX CC ftp.wipo.int/pub/published_pct_sequences

XX SQ Sequence 43 AA;

Query Match 1.3%; Score 7; DB 4; Length 43;

Best Local Similarity 100.0%; Pred. No. 1.6e+02;

Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 477 QGSLTDS 483

DB 26 QGSLTDS 32

RESULT 77

ABG23761

ID ABG23761 standard; protein; 43 AA.

AC ABG23761;

DT 18-FEB-2002 (first entry)

DE Novel human diagnostic protein #23752.

XX Human; chromosome mapping; gene mapping; gene therapy; forensics;

XX food supplement; medical imaging; diagnostic; genetic disorder.

XX Homo sapiens.

XX WO200175067-A2.

XX 11-OCT-2001.

XX 30-MAR-2001; 2001WO-US008631.

XX 31-MAR-2000; 2000US-00540217.

XX 23-AUG-2000; 2000US-00649167.

XX (HYSE-) HYSEQ INC.

XX PI Drmanac RT, Liu C, Tang YT;

XX DR WPI; 2001-639362/73.

XX DR N-PSDB; AAS87948.

XX PT New isolated polynucleotide and encoded polypeptides, useful in

XX PT diagnostics, forensics, gene mapping, identification of mutations

XX PT responsible for genetic disorders or other traits and to assess

XX PT biodiversity.

XX PS Claim 20; SEQ ID NO 54120; 103pp; English.

XX CC The invention relates to isolated polynucleotide (I) and polypeptide (II)

XX CC sequences. (I) is useful as hybridisation probes, polymerase chain

XX CC reaction (PCR) primers, oligomers, and for chromosome and gene mapping,

XX CC and in recombinant production of (II). The polynucleotides are also used

XX CC in diagnostics as expressed sequence tags for identifying expressed

XX CC genes. (I) is useful in gene therapy techniques to restore normal

XX CC activity of (II) or to treat disease states involving (II). (II) is

XX CC useful for generating antibodies against it, detecting or quantitating a

XX CC polypeptide in tissue, as molecular weight markers and as a food

XX CC supplement. (II) and its binding partners are useful in medical imaging

XX CC of sites expressing (II). (I) and (II) are useful for treating disorders

XX CC involving aberrant protein expression or biological activities. The

XX CC polypeptide and polynucleotide sequences have applications in

XX CC diagnostics, forensics, gene mapping, identification of mutations

XX CC responsible for genetic disorders or other traits to assess biodiversity

XX CC and to produce other types of data and products dependent on DNA and

XX CC amino acid sequences. ABG00010-ABG30377 represent novel human diagnostic

XX CC amino acid sequences of the invention. Note: The sequence data for this

XX CC patent did not appear in the printed specification, but was obtained in

XX CC electronic format directly from WIPO at

XX CC ftp.wipo.int/pub/published_pct_sequences

XX SQ Sequence 43 AA;

Query Match 1.3%; Score 7; DB 4; Length 43;

Best Local Similarity 100.0%; Pred. No. 1.6e+02;

Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 477 QGSLTDS 483

DB 26 QGSLTDS 32

RESULT 78

AA95964

ID AA95964 standard; protein; 44 AA.

AC AA95964;

XX 20-NOV-2000 (first entry)

DE Construct used in post-translational modification assay.

XX Coiled-coil; post-translational modification; protease; assay; assay;

XX signal modulation.

XX Synthetic.

XX Key Location/Qualifiers

XX FT Misc-difference 1 /note= "amino acid used for measuring reporter output"

XX FT Peptide 2..15 /label= Leader_Peptide

XX FT Misc-difference 16..19 /note= "GPI attachment site"

XX FT Peptide 20..44 /note= "recognition sequence; residues 20-29 are hydrophilic and residues 30-44 are hydrophobic"

XX WO200050635-A1.

XX 31-AUG-2000.
 PD 18-FEB-2000; 2000WO-GB000601.
 XX 25-FEB-1999; 99GB-00004401.
 XX (FLUO-) FLUORESCENCE LTD.
 XX Coyer J, Lightowler J;
 XX WPI; 2000-565459/52.
 DR Labeled polypeptide binding partner compositions useful for monitoring
 XX protease activity by detecting signal modulation.
 XX Example 6; Page 63; 90pp; English.
 PS The present sequence is that of a polypeptide that adopts a
 CC heteromultimeric coiled-coil conformation. It is used in an example of
 CC the invention to measure post-translational modification events which
 CC have proteolysis as an integral step. A polypeptide useful in the present
 CC invention is capable of multimerising with similar or different
 CC polypeptides. The multimer comprises a first polypeptide associated with
 CC a label and a second polypeptide, where (a) at least 1 of the
 CC polypeptides is susceptible to protease digestion, (b) association of the
 CC polypeptides to form a multimer is detectable via a signal emitted by the
 CC signal, and (c) digestion of at least 1 polypeptides results in
 CC dissociation of the multimer and modulation of the signal emitted by the
 CC label
 XX Sequence 44 AA;
 SQ
 Query Match 1.3%; Score 7; DB 3; Length 44;
 Best Local Similarity 100.0%; Pred. No. 1.6e+02;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 406 IKLLEKX 412
 DB 23 IKLLEKE 29
 |||||
 RESULT 79
 AAY76347
 ID AAY76347 standard; protein; 45 AA.
 XX
 AC AAY76347;
 XX
 DT 23-MAR-2000 (first entry)
 XX
 DE Fragment of human secreted protein encoded by gene 48.
 XX
 KW Human; secreted protein; cancer; tumour; developmental abnormality;
 KW foetal deficiency; blood disorder; immune system disorder; inflammation;
 KW autoimmune disease; allergy; Alzheimer's disease; cognitive disorder;
 KW schizophrenia; arthritis; asthma; psoriasis; sepsis; skin disorder;
 KW atherosclerosis; diabetes; cardiovascular disorder; kidney disorder;
 KW digestive disorder; endocrine disorder; infection; AIDS; leukaemia;
 KW therapy.
 XX
 KW Homo sapiens.
 CS
 XX WO9958660-A1.
 XX
 PD 18-NOV-1999.
 XX
 PF 06-MAY-1999; 99WO-US009847.
 XX
 PR 12-MAY-1998; 98US-0085093P.
 PR 12-MAY-1998; 98US-0085105P.
 PR 12-MAY-1998; 98US-0085180P.
 PR 18-MAY-1998; 98US-0085906P.

PR 18-MAY-1998; 98US-0085920P.
 PR 18-MAY-1998; 98US-0085921P.
 PR 18-MAY-1998; 98US-0085922P.
 PR 18-MAY-1998; 98US-0085923P.
 PR 18-MAY-1998; 98US-0085924P.
 PR 18-MAY-1998; 98US-0085925P.
 PR 18-MAY-1998; 98US-0085926P.
 PR 18-MAY-1998; 98US-0085928P.
 XX
 PA (HUMA-) HUMAN GENOME SCI INC.
 XX
 XX Ruben SM, Florence K, Ni J, Rosen CA, Carter KC, Moore PA;
 PI Olsen HS, Shi Y, Young PE, Wei P, Brewer LA, Soppet DR, Lafleur DW;
 PI Endress GA, Ebner R;
 XX
 DR WPI; 2000-062296/05.
 XX
 XX New isolated human genes and the secreted polypeptides they encode,
 PT useful for diagnosis and treatment of e.g. cancers, neurological
 PT disorders, immune diseases, inflammation or blood disorders.
 XX
 PS Disclosure; Page 452; 475pp; English.
 XX
 CC AAZ65250 to AAZ65350 represent 97 isolated human secreted protein genes.
 CC AAY76124 to AAY76223 are the secreted proteins encoded by the 97 human
 CC genes. This sequence represents a fragment of one of the human secreted
 CC proteins. The genes and their corresponding secreted polypeptides are
 CC useful for preventing, treating or ameliorating medical conditions, e.g.
 CC by protein or gene therapy. Also pathological conditions can be diagnosed
 CC by determining the amount of the new polypeptides in a sample or by
 CC determining the presence of mutations in the new genes. Specific uses are
 CC described for each of the 97 genes, based on which tissues they are most
 CC highly expressed in, and include developing products for the diagnosis or
 CC treatment of cancer, tumours, developmental abnormalities and foetal
 CC deficiencies, blood disorders, diseases of the immune system, autoimmune
 CC diseases, inflammation, allergies, Alzheimer's and cognitive disorders,
 CC schizophrenia, arthritis, asthma, psoriasis, sepsis, skin disorders,
 CC atherosclerosis, diabetes, cardiovascular disorders, kidney disorders,
 CC digestive/endocrine disorders, infections and AIDS. The polypeptides are
 CC also useful for identifying their binding partners. The sequences shown
 CC in AAY76224 to AAY76424 represent fragments of the secreted proteins
 XX
 SQ Sequence 45 AA;
 Query Match 1.3%; Score 7; DB 3; Length 45;
 Best Local Similarity 100.0%; Pred. No. 1.6e+02;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 12 EGAQERP 18
 DB 34 EGAQERP 40
 |||||
 RESULT 80
 AAB56241
 ID AAB56241 standard; protein; 45 AA.
 XX
 AC AAB56241;
 XX
 DT 13-MAR-2001 (first entry)
 XX
 DE Human secreted protein sequence encoded by gene 8 SEQ ID NO:335.
 XX
 KW Human; secreted protein; diagnosis; immunosuppressive; antiarthritic;
 KW antirheumatic; antiproliferative; cytostatic; cardiant; vasotropic;
 KW cerebroprotective; nootropic; neuroprotective; antibacterial; virucide;
 KW fungicide; ophthalmological; gene therapy; pathological condition;
 KW autoimmune disease; rheumatoid arthritis; hyperproliferative disorder;
 KW neoplasm; cardiovascular disorder; cardiac arrest; cerebral ischaemia;
 KW cerebrovascular disorder; angiogenesis; nervous system disorder;
 KW Alzheimer's disease; infection; ocular disorder; corneal infection;
 KW wound healing; skin aging; food additive; preservative.
 XX

OS Homo sapiens.
XX WO200070042-A1.
XX 23-NOV-2000.
XX 11-MAY-2000; 2000WO-US012788.
XX 13-MAY-1999; 99US-01340689.
XX (HUMA-) HUMAN GENOME SCI INC.
XX Rosen CA, Ruben SM, Moore PA, Young PE, Komatsoulis GA, Birse CE;
PI Duan RD, Florence KA, Soppet DR;
XX WPI; 2000-679828/66.
XX Isolated nucleic acid molecule encoding a human secreted protein is used
PT in preventing, treating or ameliorating a medical condition.
XX
XX Disclosure; Page 1010; 1065pp; English.
XX The polynucleotide sequences given in AAC99818 to AAC99977 encode the
CC human secreted proteins given in AAB56077 to AAB56362. Human secreted
CC proteins have activities based on the tissues and cells the genes are
CC expressed in. Examples of activities include: immunosuppressive;
CC antithrombotic; antirheumatic; antiproliferative; cytostatic; cardiant;
CC vasotropic; cerebroprotective; nootropic; neuroprotective; antibacterial;
CC virucide; fungicide; and ophthalmological. The human secreted
CC polynucleotides and proteins can be used to prevent, treat or ameliorate
CC a medical condition in e.g. humans, mice, rabbits, goats, horses, cats,
CC dogs, chickens or sheep. They are also used in diagnosing a pathological
CC condition or susceptibility to a pathological condition. Disorders which
CC are diagnosed or treated include autoimmune diseases e.g. rheumatoid
CC arthritis, hyperproliferative disorders e.g. neoplasms of the breast or
CC liver, cardiovascular disorders e.g. cardiac arrest, cerebrovascular
CC disorders e.g. cerebral ischaemia, angiogenesis, nervous system disorders
CC e.g. Alzheimer's disease, infections caused by bacteria, viruses and
CC fungi and ocular disorders e.g. corneal infection. The proteins can also
CC be used to aid wound healing and epithelial cell proliferation, to
CC prevent skin aging due to sunburn, to maintain organs before
CC transplantation, for supporting cell culture of primary tissues, to
CC regenerate tissues and in chemotaxis. The proteins can also be used as a
CC food additive or preservative to increase or decrease storage
CC capabilities. AAC99809 to AAC99817 and AAB56076 represent sequences used
CC in the exemplification of the present invention
XX
XX Sequence 45 AA;
XX
XX Query Match 1.3%; Score 7; DB 3; Length 45;
XX Best Local Similarity 100.0%; Pred. No. 1.6e+02;
XX Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
XX
XX 12 EGAQERP 18
XX |||||
XX 34 EGAQERP 40
XX
XX RESULT 81
XX ADEL1987
XX ID ADEL1987 standard; protein; 45 AA.
XX
XX ADEL1987;
XX AC
XX 29-JAN-2004 (first entry)
XX
XX Human secreted polypeptide #241.
XX
XX Secreted protein; cancer; liver disorder; hepatitis; neural disorder;
XX Alzheimer's disease; human.
XX Synthetic.
XX Homo sapiens.

XX US2003100051-A1.
XX 29-MAY-2003.
XX 10-SEP-2001; 2001US-00948783.
XX 12-MAY-1998; 98US-0085093P.
XX 12-MAY-1998; 98US-0085094P.
XX 12-MAY-1998; 98US-0085105P.
XX 12-MAY-1998; 98US-0085180P.
XX 18-MAY-1998; 98US-0085906P.
XX 18-MAY-1998; 98US-0085920P.
XX 18-MAY-1998; 98US-0085921P.
XX 18-MAY-1998; 98US-0085922P.
XX 18-MAY-1998; 98US-0085923P.
XX 18-MAY-1998; 98US-0085924P.
XX 18-MAY-1998; 98US-0085925P.
XX 18-MAY-1998; 98US-0085927P.
XX 18-MAY-1998; 98US-0085928P.
XX 06-MAY-1999; 99WO-US009847.
XX 10-NOV-1999; 99US-00437858.
XX 11-SEP-2000; 2000US-0231846P.
XX 23-JUN-2001; 2001US-00892877.
XX (RUBE/) RUBEN S M.
XX (FLO/) FLORENCE K A.
XX (NIJ/) NI J.
XX (ROSE/) ROSEN C A.
XX (CART/) CARTER K C.
XX (MOOR/) MOORE P A.
XX (OLSE/) OLSEN H S.
XX (SHI/) SHI Y.
XX (YOUN/) YOUNG P E.
XX (WEI/) WEI Y.
XX (BREW/) BREWER L A.
XX (SOPP/) SOPPET D R.
XX (LAF/) LAFLEUR D W.
XX (ENDR/) ENDRESS G A.
XX (BERN/) BERNER R.
XX (BIRS/) BIRSE C E.
XX Ruben SM, Florence KA, Ni J, Rosen CA, Carter KC, Moore PA;
PI Olsen HS, Shi Y, Young PE, Wei Y, Brewer LA, Soppet DR, Lafleur DW;
PI Endress GA, Ebner R, Birse CE;
XX WPI, 2003-801210/75.
XX
XX New nucleic acid molecule, useful for preparing a medicament for
PT preventing, treating or ameliorating a medical condition e.g. Cancer,
PT liver disorders or neural disorders.
XX
XX Claim 11; SEQ ID NO 359; 453pp; English.
XX
XX The invention relates to human secreted polypeptides and the
CC polynucleotides encoding them. The sequences are useful for preparing
CC medicaments for preventing, treating or ameliorating medical conditions
CC e.g., cancer, liver disorders such as hepatitis or neural disorders such
CC as Alzheimer's disease. This sequence represents a human secreted
CC polypeptide of the invention.
XX
XX Sequence 45 AA;
XX
XX Query Match 1.3%; Score 7; DB 7; Length 45;
XX Best Local Similarity 100.0%; Pred. No. 1.6e+02;
XX Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
XX
XX 12 EGAQERP 18
XX |||||
XX 34 EGAQERP 40
XX
XX RESULT 82

PN W0620953-A2.
XX 11-JUL-1996.
XX 22-DEC-1995; 95WO-US016854.
XX 30-DEC-1994; 94US-00366479.
XX (UYNC-) UNIV NORTH CAROLINA.
XX Lemon SM, Erickson BW, Wang JG, Rozzelle J;
XX WPI; 1996-333940/33.
XX Synthetic peptide with Delta hepatitis virus antigenic activity - useful
XX to produce and detect viral antibodies in a mammal and for immunisation.
XX Claim 1; Page 42; 50pp; English.
XX The sequences given in AAW03428-40 represent synthetic peptides with
XX hepatitis delta virus (HDV) antigenic activity. The peptide, delta12-
XX 60(Y) was designed to mimic the coiled-coil dimerisation domain of the
XX HDV protein. Circular dichroism spectroscopy has demonstrated that this
XX protein has a strong concentration-dependent tendency for alpha-helical
XX coiled-coil complexes, with a Tm in excess of 80 deg. C. This peptide
XX demonstrates exceptionally strong and broadly reactive antigenic activity
XX and expresses conformational epitopes of HDV. It is thought to self-
XX assemble into a multimeric structure composed of four or more peptide
XX chains. The sequences given in W032429-40 represent analogs of delta12-
XX 60(Y). These peptides may be used to detect antibodies to HDV in a
XX mammal, and to produce neutralising antibodies to immunise animals
XX against HDV
XX
SQ Sequence 50 AA;
Query Match 1.3%; Score 7; DB 2; Length 50;
Best Local Similarity 100.0%; Pred. No. 1.8e+02; Indels 0; Gaps 0;
Matches 7; Conservative 0; Mismatches 0;
QY 404 KKIKKLE 410
DB 28 KKIKKLE 34
RESULT 84
AAW03429
ID AAW03429 standard; peptide; 50 AA.
XX AAW03429;
XX AC AAW03429;
XX 14-FEB-1997 (first entry)
XX HDV antigenic peptide analog, delta12-60(Y)/S22C.
XX Hepatitis delta virus; HDV; antigen; delta12-60(Y); alpha-helix;
XX coiled-coil dimerisation domain; epitope; multimer.
XX Synthetic.
XX W0620953-A2.
XX 11-JUL-1996.
XX 22-DEC-1995; 95WO-US016854.
XX 30-DEC-1994; 94US-00366479.
XX (UYNC-) UNIV NORTH CAROLINA.
XX Lemon SM, Erickson BW, Wang JG, Rozzelle J;
XX WPI; 1996-333940/33.
XX

AAW03431
ID AAW03431 standard; peptide; 50 AA.
XX AAW03431;
XX 14-FEB-1997 (first entry)
XX HDV antigenic peptide analog, delta12-60(Y)/Fr.
XX Hepatitis delta virus; HDV; antigen; delta12-60(Y); alpha-helix;
XX coiled-coil dimerisation domain; epitope; multimer.
XX Synthetic.
XX W0620953-A2.
XX 11-JUL-1996.
XX 22-DEC-1995; 95WO-US016854.
XX 30-DEC-1994; 94US-00366479.
XX (UYNC-) UNIV NORTH CAROLINA.
XX Lemon SM, Erickson BW, Wang JG, Rozzelle J;
XX WPI; 1996-333940/33.
XX Synthetic peptide with Delta hepatitis virus antigenic activity - useful
XX to produce and detect viral antibodies in a mammal and for immunisation.
XX Claim 1; Page 34; 50pp; English.
XX The sequences given in AAW03428-40 represent synthetic peptides with
XX hepatitis delta virus (HDV) antigenic activity. The peptide, delta12-
XX 60(Y) was designed to mimic the coiled-coil dimerisation domain of the
XX HDV protein. Circular dichroism spectroscopy has demonstrated that this
XX protein has a strong concentration-dependent tendency to form alpha-
XX helical coiled-coil complexes, with a Tm in excess of 80 deg. C. This
XX peptide demonstrates exceptionally strong and broadly reactive antigenic
XX activity and expresses conformational epitopes of HDV. It is thought to
XX self-assemble into a multimeric structure composed of four or more
XX peptide chains. The sequences given in W032429-40 represent analogs of
XX delta12-60(Y). These peptides may be used to detect antibodies to HDV in
XX a mammal, and to produce neutralising antibodies to immunise animals
XX against HDV
XX
SQ Sequence 50 AA;
Query Match 1.3%; Score 7; DB 2; Length 50;
Best Local Similarity 100.0%; Pred. No. 1.8e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 404 KKIKKLE 410
DB 28 KKIKKLE 34
RESULT 83
AAW03428
ID AAW03428 standard; peptide; 50 AA.
XX AAW03428;
XX 14-FEB-1997 (first entry)
XX Delta hepatitis virus derived peptide, delta12-60(Y).
XX Hepatitis delta virus; HDV; antigen; delta12-60(Y); alpha-helix;
XX coiled-coil dimerisation domain; epitope; multimer.
XX Synthetic.
XX

PT Synthetic peptide with Delta hepatitis virus antigenic activity - useful
PT to produce and detect viral antibodies in a mammal and for immunisation.
XX
PS Claim 1; Page 33; 50pp; English.
XX
CC The sequences given in AAW03428-40 represent synthetic peptides with
CC hepatitis delta virus (HDV) antigenic activity. The peptide, delta12-
CC 60(Y) was designed to mimic the coiled-coil dimerisation domain of the
CC HDV protein. Circular dichroism spectroscopy has demonstrated that this
CC protein has a strong concentration-dependent tendency to form alpha-
CC helical coiled-coil complexes, with a Tm in excess of 80 deg. C. This
CC peptide demonstrates exceptionally strong and broadly reactive antigenic
CC activity and expresses conformational epitopes of HDV. It is thought to
CC self-assemble into a multimeric structure composed of four or more
CC peptide chains. The sequences given in W032429-40 represent analogs of
CC delta12-60(Y). These peptides may be used to detect antibodies to HDV in
CC a mammal, and to produce neutralising antibodies to immunise animals
CC against HDV
XX
SQ Sequence 50 AA;
XX
Query Match 1.3%; Score 7; DB 2; Length 50;
Best Local Similarity 100.0%; Pred. No. 1.8e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 404 KKIKKLE 410
DB 28 KKIKKLE 34
RESULT 85
AAW03438
ID AAW03438 standard; peptide; 50 AA.
AC AAW03438;
XX
DT 14-FEB-1997 (first entry)
XX
DE HDV antigenic peptide analog, delta12-60(Y)/Fe.
XX
KW Hepatitis delta virus; HDV; antigen; delta12-60(Y); alpha-helix;
KW coiled-coil dimerisation domain; epitope; multimer.
XX
OS Synthetic.
XX
FN W09620953-A2.
XX
PD 11-JUL-1996.
XX
PF 22-DEC-1995; 95WO-US016854.
XX
PR 30-DEC-1994; 94US-00366479.
XX
PA (UYNCL) UNIV NORTH CAROLINA.
XX
PI Lemon SM, Erickson BW, Wang JG, Rozzelle J;
XX
DR WPI; 1996-333940/33.
XX
PT Synthetic peptide with Delta hepatitis virus antigenic activity - useful
PT to produce and detect viral antibodies in a mammal and for immunisation.
XX
PS Claim 1; Page 37; 50pp; English.
XX
CC The sequences given in AAW03428-40 represent synthetic peptides with
CC hepatitis delta virus (HDV) antigenic activity. The peptide, delta12-
CC 60(Y) was designed to mimic the coiled-coil dimerisation domain of the
CC HDV protein. Circular dichroism spectroscopy has demonstrated that this
CC protein has a strong concentration-dependent tendency to form alpha-
CC helical coiled-coil complexes, with a Tm in excess of 80 deg. C. This
CC peptide demonstrates exceptionally strong and broadly reactive antigenic
CC activity and expresses conformational epitopes of HDV. It is thought to
CC self-assemble into a multimeric structure composed of four or more

CC peptide chains. The sequences given in W032429-40 represent analogs of
CC delta12-60(Y). These peptides may be used to detect antibodies to HDV in
CC a mammal, and to produce neutralising antibodies to immunise animals
CC against HDV
XX
SQ Sequence 50 AA;
XX
Query Match 1.3%; Score 7; DB 2; Length 50;
Best Local Similarity 100.0%; Pred. No. 1.8e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 404 KKIKKLE 410
DB 28 KKIKKLE 34
RESULT 86
AAW03439
ID AAW03439 standard; peptide; 50 AA.
AC AAW03439;
XX
DT 14-FEB-1997 (first entry)
XX
DE HDV antigenic peptide analog, delta12-60(Y)/Ta.
XX
KW Hepatitis delta virus; HDV; antigen; delta12-60(Y); alpha-helix;
KW coiled-coil dimerisation domain; epitope; multimer.
XX
OS Synthetic.
XX
FN W09620953-A2.
XX
PD 11-JUL-1996.
XX
PF 22-DEC-1995; 95WO-US016854.
XX
PR 30-DEC-1994; 94US-00366479.
XX
PA (UYNCL) UNIV NORTH CAROLINA.
XX
PI Lemon SM, Erickson BW, Wang JG, Rozzelle J;
XX
DR WPI; 1996-333940/33.
XX
PT Synthetic peptide with Delta hepatitis virus antigenic activity - useful
PT to produce and detect viral antibodies in a mammal and for immunisation.
XX
PS Claim 1; Page 38; 50pp; English.
XX
CC The sequences given in AAW03428-40 represent synthetic peptides with
CC hepatitis delta virus (HDV) antigenic activity. The peptide, delta12-
CC 60(Y) was designed to mimic the coiled-coil dimerisation domain of the
CC HDV protein. Circular dichroism spectroscopy has demonstrated that this
CC protein has a strong concentration-dependent tendency to form alpha-
CC helical coiled-coil complexes, with a Tm in excess of 80 deg. C. This
CC peptide demonstrates exceptionally strong and broadly reactive antigenic
CC activity and expresses conformational epitopes of HDV. It is thought to
CC self-assemble into a multimeric structure composed of four or more
CC peptide chains. The sequences given in W032429-40 represent analogs of
CC delta12-60(Y). These peptides may be used to detect antibodies to HDV in
CC a mammal, and to produce neutralising antibodies to immunise animals
CC against HDV
XX
SQ Sequence 50 AA;
XX
Query Match 1.3%; Score 7; DB 2; Length 50;
Best Local Similarity 100.0%; Pred. No. 1.8e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 404 KKIKKLE 410
DB 28 KKIKKLE 34

RESULT 87
AAW03437
ID AAW03437 standard; peptide; 50 AA.

AC AAW03437;
XX
DT 14-FEB-1997 (first entry)
XX
DE HDV antigenic peptide analog, delta12-60(Y)/Na.
XX
KW Hepatitis delta virus; HDV; antigen; delta12-60(Y); alpha-helix;
KW coiled-coil dimerisation domain; epitope; multimer.
XX
OS Synthetic.

XX
XX W09620953-A2.
XX
XX 11-JUL-1996.
XX
XX 22-DEC-1995; 95WO-US016854.
XX
XX 30-DEC-1994; 94US-00366479.
XX
XX (UYNC-) UNIV NORTH CAROLINA.

XX
PI Lemon SM, Erickson BW, Wang JG, Rozzelle J;
XX
XX WPI; 1996-333940/33.

XX
XX Synthetic peptide with Delta hepatitis virus antigenic activity - useful
XX to produce and detect viral antibodies in a mammal and for immunisation.
XX
XX Claim 1; Page 37; 50pp; English.

XX
CC The sequences given in AAW03428-40 represent synthetic peptides with
CC hepatitis delta virus (HDV) antigenic activity. The peptide, delta12-
CC 60(Y) was designed to mimic the coiled-coil dimerisation domain of the
CC HDV protein. Circular dichroism spectroscopy has demonstrated that this
CC protein has a strong concentration-dependent tendency to form alpha-
CC helical coiled-coil complexes, with a Tm in excess of 80 deg. C. This
CC peptide demonstrates exceptionally strong and broadly reactive antigenic
CC activity and expresses conformational epitopes of HDV. It is thought to
CC self-assemble into a multimeric structure composed of four or more
CC peptide chains. The sequences given in W032429-40 represent analogs of
CC delta12-60(Y). These peptides may be used to detect antibodies to HDV in
CC a mammal, and to produce neutralising antibodies to immunise animals
CC against HDV

XX
SQ Sequence 50 AA;

Query Match 1.3%; Score 7; DB 2; Length 50;
Best Local Similarity 100.0%; Pred. No. 1.8e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 404 KKIKKLE 410
DB 28 KKIKKLE 34

RESULT 88
AAW03434
ID AAW03434 standard; peptide; 50 AA.

AC AAW03434;
XX
DT 14-FEB-1997 (first entry)
XX
DE HDV antigenic peptide analog, delta12-60(Y)/Jal.
XX
KW Hepatitis delta virus; HDV; antigen; delta12-60(Y); alpha-helix;
KW coiled-coil dimerisation domain; epitope; multimer.

XX
OS Synthetic.
XX
XX W09620953-A2.
XX
XX 11-JUL-1996.

XX
XX 22-DEC-1995; 95WO-US016854.
XX
XX 30-DEC-1994; 94US-00366479.
XX
XX (UYNC-) UNIV NORTH CAROLINA.

XX
PI Lemon SM, Erickson BW, Wang JG, Rozzelle J;
XX
XX WPI; 1996-333940/33.

XX
XX Synthetic peptide with Delta hepatitis virus antigenic activity - useful
XX to produce and detect viral antibodies in a mammal and for immunisation.
XX
XX Claim 1; Page 35; 50pp; English.

XX
CC The sequences given in AAW03428-40 represent synthetic peptides with
CC hepatitis delta virus (HDV) antigenic activity. The peptide, delta12-
CC 60(Y) was designed to mimic the coiled-coil dimerisation domain of the
CC HDV protein. Circular dichroism spectroscopy has demonstrated that this
CC protein has a strong concentration-dependent tendency to form alpha-
CC helical coiled-coil complexes, with a Tm in excess of 80 deg. C. This
CC peptide demonstrates exceptionally strong and broadly reactive antigenic
CC activity and expresses conformational epitopes of HDV. It is thought to
CC self-assemble into a multimeric structure composed of four or more
CC peptide chains. The sequences given in W032429-40 represent analogs of
CC delta12-60(Y). These peptides may be used to detect antibodies to HDV in
CC a mammal, and to produce neutralising antibodies to immunise animals
CC against HDV

XX
SQ Sequence 50 AA;

Query Match 1.3%; Score 7; DB 2; Length 50;
Best Local Similarity 100.0%; Pred. No. 1.8e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 404 KKIKKLE 410
DB 28 KKIKKLE 34

RESULT 89
AAW03436
ID AAW03436 standard; peptide; 50 AA.

XX
XX AAW03436;
XX
DT 14-FEB-1997 (first entry)
XX
DE HDV antigenic peptide analog, delta12-60(Y)/Le.
XX
KW Hepatitis delta virus; HDV; antigen; delta12-60(Y); alpha-helix;
KW coiled-coil dimerisation domain; epitope; multimer.

XX
OS Synthetic.
XX
XX W09620953-A2.
XX
XX 11-JUL-1996.

XX
XX 22-DEC-1995; 95WO-US016854.
XX
XX 30-DEC-1994; 94US-00366479.
XX
XX (UYNC-) UNIV NORTH CAROLINA.

XX
PI Lemon SM, Erickson BW, Wang JG, Rozzelle J;

XX WPI; 1996-333940/33.
XX
XX Synthetic peptide with Delta hepatitis virus antigenic activity - useful
PT to produce and detect viral antibodies in a mammal and for immunisation.
XX
XX
PS Claim 1; Page 36; 50pp; English.
XX
XX The sequences given in AA03428-40 represent synthetic peptides with
CC hepatitis delta virus (HDV) antigenic activity. The peptide, delta12-
CC 60(Y) was designed to mimic the coiled-coil dimerisation domain of the
CC HDV protein. Circular dichroism spectroscopy has demonstrated that this
CC protein has a strong concentration-dependent tendency to form alpha-
CC helical coiled-coil complexes, with a Tm in excess of 80 deg. C. This
CC peptide demonstrates exceptionally strong and broadly reactive antigenic
CC activity and expresses conformational epitopes of HDV. It is thought to
CC self-assemble into a multimeric structure composed of four or more
CC peptide chains. The sequences given in W032429-40 represent analogs of
CC delta12-60(Y). These peptides may be used to detect antibodies to HDV in
CC a mammal, and to produce neutralising antibodies to immunise animals,
XX against HDV
XX
SQ Sequence 50 AA;
Query Match 1.3%; Score 7; DB 2; Length 50;
Best Local Similarity 100.0%; Pred. No. 1.8e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 404 KKIKKLE 410
Db 28 KKIKKLE 34
RESULT 90
ABG10316
ID ABG10316 standard; peptide; 50 AA.
XX
XX ABG10316;
AC
XX
XX 13-FEB-2002 (first entry)
DT
XX
XX Novel human diagnostic protein #10307.
DE
XX Human; chromosome mapping; gene mapping; gene therapy; forensic;
KW food supplement; medical imaging; diagnostic; genetic disorder.
KW
XX Homo sapiens.
OS
XX
XX WO200175067-A2.
PN
XX
XX 11-OCT-2001.
PD
XX
XX 30-MAR-2001; 2001WO-US008631.
PF
XX
XX 31-MAR-2000; 2000US-00540217.
PR
XX
XX 23-AUG-2000; 2000US-00649167.
PR
XX
XX (HYSE-) HYSEQ INC.
PA
XX
XX Drmanac RT, Liu C, Tang YT;
PI
XX
XX WPI; 2001-639362/73.
DR
XX
XX N-PSDB; AAS74503.
DR
XX
XX New isolated polynucleotide and encoded polypeptides, useful in
PT diagnostics, forensics, gene mapping, identification of mutations
PT responsible for genetic disorders or other traits and to assess
PT biodiversity.
XX
XX Claim 20; SEQ ID NO 40675; 103pp; English.
PS
XX
XX The invention relates to isolated polynucleotide (I) and polypeptide (II)
CC sequences. (I) is useful as hybridisation probes, polymerase chain
CC reaction (PCR) primers, oligomers, and for chromosome and gene mapping.
CC and in recombinant production of (II). The polynucleotides are also used
CC in diagnostics as expressed sequence tags for identifying expressed
CC genes. (I) is useful in gene therapy techniques to restore normal
CC activity of (II) or to treat disease states involving (II). (II) is
CC useful for generating antibodies against it, detecting or quantitating a
CC polypeptide in tissue, as molecular weight markers and as a food
CC supplement. (II) and its binding partners are useful in medical imaging
CC of sites expressing (II). (I) and (II) are useful for treating disorders
CC involving aberrant protein expression or biological activity. The
CC polypeptide and polynucleotide sequences have applications in
CC diagnostics, forensics, gene mapping, identification of mutations
CC responsible for genetic disorders or other traits to assess biodiversity
CC and to produce other types of data and products dependent on DNA and

XX WPI; 1996-333940/33.
XX
XX Synthetic peptide with Delta hepatitis virus antigenic activity - useful
PT to produce and detect viral antibodies in a mammal and for immunisation.
XX
XX
PS Claim 1; Page 36; 50pp; English.
XX
XX The sequences given in AA03428-40 represent synthetic peptides with
CC hepatitis delta virus (HDV) antigenic activity. The peptide, delta12-
CC 60(Y) was designed to mimic the coiled-coil dimerisation domain of the
CC HDV protein. Circular dichroism spectroscopy has demonstrated that this
CC protein has a strong concentration-dependent tendency to form alpha-
CC helical coiled-coil complexes, with a Tm in excess of 80 deg. C. This
CC peptide demonstrates exceptionally strong and broadly reactive antigenic
CC activity and expresses conformational epitopes of HDV. It is thought to
CC self-assemble into a multimeric structure composed of four or more
CC peptide chains. The sequences given in W032429-40 represent analogs of
CC delta12-60(Y). These peptides may be used to detect antibodies to HDV in
CC a mammal, and to produce neutralising antibodies to immunise animals,
XX against HDV
XX
SQ Sequence 50 AA;
Query Match 1.3%; Score 7; DB 2; Length 50;
Best Local Similarity 100.0%; Pred. No. 1.8e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 404 KKIKKLE 410
Db 28 KKIKKLE 34
RESULT 90
AAW03440
ID AAW03440 standard; peptide; 50 AA.
XX
XX AAW03440;
AC
XX
XX 14-FEB-1997 (first entry)
DT
XX
XX HDV antigenic peptide analog, delta12-60(Y)-Cons.
DE
XX
XX Hepatitis delta virus; HDV; antigen; delta12-60(Y); alpha-helix;
KW coiled-coil dimerisation domain; epitope; multimer.
KW
XX Synthetic.
OS
XX
XX WO9620953-A2.
PN
XX
XX 11-JUL-1996.
PD
XX
XX 22-DEC-1995; 95WO-US016854.
PF
XX
XX 30-DEC-1994; 94US-00366479.
PR
XX
XX (UYNC-) UNIV NORTH CAROLINA.
PA
XX
XX Lemon SM, Erickson BW, Wang JG, Rozzelle J;
PI
XX
XX WPI; 1996-333940/33.
DR
XX
XX Synthetic peptide with Delta hepatitis virus antigenic activity - useful
PT to produce and detect viral antibodies in a mammal and for immunisation.
PT
XX
XX Claim 1; Page 38; 50pp; English.
PS
XX
XX The sequences given in AA03428-40 represent synthetic peptides with
CC hepatitis delta virus (HDV) antigenic activity. The peptide, delta12-
CC 60(Y) was designed to mimic the coiled-coil dimerisation domain of the
CC HDV protein. Circular dichroism spectroscopy has demonstrated that this
CC protein has a strong concentration-dependent tendency to form alpha-
CC helical coiled-coil complexes, with a Tm in excess of 80 deg. C. This

CC amino acid sequences. ABG00010-ABG30377 represent novel human diagnostic
 CC amino acid sequences of the invention. Note: The sequence data for this
 CC patent did not appear in the printed specification, but was obtained in
 CC electronic format directly from WIPO at
 CC ftp.wipo.int/pub/published_pct_sequences
 CC
 XX
 SQ Sequence 51 AA;

Query Match 1.3%; Score 7; DB 4; Length 51;
 Best Local Similarity 100.0%; Pred. No. 1.8e+02;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 247 ARBEEK 253
 Db 14 ARBEEK 20

RESULT 92
 ABG23757
 ID ABG23757 standard; protein; 52 AA.

XX AC ABG23757;

XX DT 18-FEB-2002 (first entry)

XX DE Novel human diagnostic protein #23748.

XX KW Human; chromosome mapping; gene mapping; gene therapy; forensic;
 KW food supplement; medical imaging; diagnostic; genetic disorder.

XX OS Homo sapiens.

XX PN WO200175067-A2.

XX PD 11-OCT-2001.

XX PF 30-MAR-2001; 2001WO-US008631.

XX PR 31-MAR-2000; 2000US-00540217.

XX PT 23-AUG-2000; 2000US-00649167.

XX PA (HYSE-) HYSEQ INC.

XX PI Drmanac RT, Liu C, Tang YT;

XX DR WPI; 2001-639362/73.

XX DR N-PSDB; AAS87944.

XX PT New isolated polynucleotide and encoded polypeptides, useful in
 PT diagnostics, forensics, gene mapping, identification of mutations
 PT responsible for genetic disorders or other traits and to assess
 PT biodiversity.

XX PS Claim 20; SEQ ID NO 54116; 103pp; English.

XX CC The invention relates to isolated polynucleotide (I) and polypeptide (II)
 CC sequences. (I) is useful as hybridisation probes, polymerase chain
 CC reaction (PCR) primers, oligomers, and for chromosome and gene mapping,
 CC and in recombinant production of (II). The polynucleotides are also used
 CC in diagnostics as expressed sequence tags for identifying expressed
 CC genes. (II) is useful in gene therapy techniques to restore normal
 CC activity of (II) or to treat disease states involving (II). (II) is
 CC useful for generating antibodies against it, detecting or quantitating a
 CC polypeptide in tissue, as molecular weight markers and as a food
 CC supplement. (II) and its binding partners are useful in medical imaging
 CC of sites expressing (II). (I) and (II) are useful for treating disorders
 CC involving aberrant protein expression or biological activity. The
 CC polypeptide and polynucleotide sequences have applications in
 CC diagnostics, forensics, gene mapping, identification of mutations
 CC responsible for genetic disorders or other traits to assess biodiversity
 CC and to produce other types of data and products dependent on DNA and
 CC amino acid sequences. ABG00010-ABG30377 represent novel human diagnostic
 CC amino acid sequences of the invention. Note: The sequence data for this

CC patent did not appear in the printed specification, but was obtained in
 CC electronic format directly from WIPO at
 CC ftp.wipo.int/pub/published_pct_sequences
 CC
 XX
 SQ Sequence 52 AA;

Query Match 1.3%; Score 7; DB 4; Length 52;
 Best Local Similarity 100.0%; Pred. No. 1.8e+02;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 477 QGSLTDS 483
 Db 36 QGSLTDS 42

RESULT 93

AAW03441

ID AAW03441 standard; peptide; 57 AA.

XX AC AAW03441;

XX DT 16-OCT-2003 (revised)

XX DT 14-FEB-1997 (first entry)

XX DE HDV antigen fragment, residues 9-65.

XX KW Hepatitis delta virus; HDV; antigen; delta12-60(V); alpha-helix;
 KW coiled-coil dimerisation domain; epitope; multimer.

XX OS Hepatitis D virus.

XX PH Key Location/Qualifiers

XX PT Peptide 4..16

XX PT Peptide /label= Segment A

XX PT Peptide /note= "Corresponds to HDag residues 12-24"

XX FT Peptide 17..41

XX FT Peptide /label= Segment B

XX FT Peptide /note= "Corresponds to HDag residues 25-49"

XX FT Peptide 42..52

XX FT Peptide /label= Segment C

XX FT Peptide /note= "Corresponds to HDag residues 50-60"

XX PN WO9620953-A2.

XX PD 11-JUL-1996.

XX PF 22-DEC-1995; 95WO-US016854.

XX PR 30-DEC-1994; 94US-00366479.

XX PA (UTNC-) UNIV NORTH CAROLINA.

XX PI Lemon SM, Erickson BW, Wang JG, Rozzelle J;

XX DR WPI; 1996-333940/33.

XX PT Synthetic peptide with Delta hepatitis virus antigenic activity - useful
 PT to produce and detect viral antibodies in a mammal and for immunisation.

XX PS Example 10; Page 29; 50pp; English.

XX CC This sequence represents residues 9-65 of the hepatitis delta virus (HDV)
 CC antigen (HDag) sequence. This sequence was used in the design of the
 CC sequences given in AAW03441-47. These sequences represent synthetic
 CC peptides with hepatitis delta virus (HDV) antigenic activity. They were
 CC used to raise murine monoclonal antibodies which may be used to immunise
 CC animals against HDV, and to detect anti-HDV antibodies. (Updated on 16-
 CC OCT-2003 to standardise OS field)

XX SQ Sequence 57 AA;

Query Match 1.3%; Score 7; DB 2; Length 57;
 Best Local Similarity 100.0%; Pred. No. 2e+02;

Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 404 KKIKKLE 410
| | | | |
Db 31 KKIKKLE 37

RESULT 94
ABG27967
ID ABG27967 standard; protein; 57 AA.
XX AC ABG27967;
XX DT 18-FEB-2002 (first entry)
XX DE Novel human diagnostic protein #27958.
XX KW Human; chromosome mapping; gene mapping; gene therapy; forensic;
XX XW food supplement; medical imaging; diagnostic; genetic disorder.
XX OS Homo sapiens.
XX PN WO200175067-A2.
XX PD 11-OCT-2001.
XX PF 30-MAR-2001; 2001WO-US008631.
XX PR 31-MAR-2000; 2000US-00540217.
XX PR 23-AUG-2000; 2000US-00649167.
XX (HYSE-) HYSEQ INC.
XX PA Drmanac RT, Liu C, Tang YT;
XX PI WPI; 2001-639362/73.
XX DR N-PSDB; AAS92154.
XX New isolated polynucleotide and encoded polypeptides, useful in
PT diagnostics, forensics, gene mapping, identification of mutations
PT responsible for genetic disorders or other traits and to assess
PT biodiversity.
XX Claim 20; SEQ ID NO 58325; 103pp; English.

Query Match 1.3%; Score 7; DB 4; Length 57;
Best Local Similarity 100.0%; Pred. No. 2e+02; Indels 0; Gaps 0;
Matches 7; Conservative 0; Mismatches 0;

Qy 404 KKIKKLE 410
| | | | |
Db 31 KKIKKLE 37

RESULT 94
ABG27967
ID ABG27967 standard; protein; 57 AA.
XX AC ABG27967;
XX DT 18-FEB-2002 (first entry)
XX DE Novel human diagnostic protein #27958.
XX KW Human; chromosome mapping; gene mapping; gene therapy; forensic;
XX XW food supplement; medical imaging; diagnostic; genetic disorder.
XX OS Homo sapiens.
XX PN WO200175067-A2.
XX PD 11-OCT-2001.
XX PF 30-MAR-2001; 2001WO-US008631.
XX PR 31-MAR-2000; 2000US-00540217.
XX PR 23-AUG-2000; 2000US-00649167.
XX (HYSE-) HYSEQ INC.
XX PA Drmanac RT, Liu C, Tang YT;
XX PI WPI; 2001-639362/73.
XX DR N-PSDB; AAS92154.
XX New isolated polynucleotide and encoded polypeptides, useful in
PT diagnostics, forensics, gene mapping, identification of mutations
PT responsible for genetic disorders or other traits and to assess
PT biodiversity.
XX Claim 20; SEQ ID NO 58325; 103pp; English.

Query Match 1.3%; Score 7; DB 4; Length 57;
Best Local Similarity 100.0%; Pred. No. 2e+02; Indels 0; Gaps 0;
Matches 7; Conservative 0; Mismatches 0;

Qy 477 QGSLTDS 483
| | | | |
Db 40 QGSLTDS 46

RESULT 95
ADC00699
ID ADC00699 standard; protein; 60 AA.
XX AC ADC00699;
XX DT 04-DEC-2003 (first entry)
XX DE Enterohaemorrhagic E. coli O157:H7-specific protein SEQ ID NO: 744.
XX DS enterohaemorrhagic; anti-bacterial.
XX KW Escherichia coli; O157:H7.
XX OS JP2002355074-A.
XX PN 10-DEC-2002.
XX PD 24-JAN-2002; 2002JP-00015959.
XX PF 24-JAN-2001; 2001JP-00112010.
XX PR (UVTS-) UNIV TSUKUBA.
XX PA WPI; 2003-451640/43.
XX PT Enterohaemorrhagic Escherichia coli O157:H7-specific nucleic acid molecule
XX PT and a polypeptide and its use, a polypeptide, a vector and a host cell.
XX PS Claim 3; SEQ ID NO 744; 2067pp; Japanese.
XX CC The invention relates to a novel enterohaemorrhagic Escherichia coli
XX O157:H7-specific nucleic acid molecule. A polynucleotide of the invention
XX has anti-bacterial activity. The polypeptide can be used in detection
XX and/or treatment of O157:H7 infection. The nucleotide sequence of the
XX genome of Enterohaemorrhagic E. coli O157:H7 was determined. The present
XX sequence represents an E. coli O157:H7-specific polypeptide of the
XX invention.
XX SQ Sequence 60 AA;

Query Match 1.3%; Score 7; DB 7; Length 60;
Best Local Similarity 100.0%; Pred. No. 2.1e+02; Indels 0; Gaps 0;
Matches 7; Conservative 0; Mismatches 0;

Qy 494 AQAPSSP 500
| | | | |
Db 48 AQAPSSP 54

RESULT 96
ABG02022
ID ABG02022 standard; protein; 70 AA.
XX AC ABG02022;
XX DT 13-FEB-2002 (first entry)
XX DE Novel human diagnostic protein #2013.
XX KW Human; chromosome mapping; gene mapping; gene therapy; forensic;
XX XW food supplement; medical imaging; diagnostic; genetic disorder.
XX OS Homo sapiens.
XX PN WO200175067-A2.
XX PD 11-OCT-2001.


```
XX 28-FEB-2000; 2000US-00515126.
PR 18-MAY-2000; 2000US-00577409.
XX (HYSE-) HYSEQ INC.
XX Tang YT, Liu C, Drmanac RT;
XX WPI; 2001-514838/56.
XX DR N-PSDB; AAI81397.
XX Isolated nucleic acids and polypeptides, useful for preventing diagnosing
PT and treating e.g. leukemia, inflammation and immune disorders.
XX Claim 20; SEQ ID NO 15358; 1399pp + Sequence Listing; English.
XX The invention relates to human polynucleotides (AAI79941-AAI93941) and
CC the encoded proteins (AAO00010-AAO13910) that exhibit activity elating to
CC cytokine, cell proliferation or cell differentiation or which may induce
CC production of other cytokines in other cell populations. The
CC polynucleotides and polypeptides are useful in gene therapy, vaccines or
CC peptide therapy. The polypeptides have various cytokine-like activities,
CC e.g. stem cell growth factor activity, haematopoiesis regulating
CC activity, tissue growth factor activity, immunomodulatory activity and
CC activin/inhibin activity and may be useful in the diagnosis and/or
CC treatment of cancer, leukaemia, nervous system disorders, arthritis and
CC inflammation. Note: The sequence data for this patent did not form part
CC of the printed specification, but was obtained in electronic format
CC directly from WIPO at ftp.wipo.int/pub/published_pct_sequences
XX
XX SQ Sequence 73 AA;
Query Match 1.3%; Score 7; DB 4; Length 73;
Best Local Similarity 100.0%; Pred. No. 2.4e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 14 AQERPSQ 20
Db 58 AQERPSQ 64
|||||
RESULT 99
ABG26987
ID ABG26987 standard; protein; 75 AA.
XX AC ABG26987;
XX DT 18-FEB-2002 (first entry)
XX DE Novel human diagnostic protein #26978.
XX KW Human; chromosome mapping; gene mapping; gene therapy; forensic;
XX KW food supplement; medical imaging; diagnostic; genetic disorder.
XX OS Homo sapiens.
XX PN WO200175067-A2.
XX PD 11-OCT-2001.
XX PF 30-MAR-2001; 2001WO-US008631.
XX PR 31-MAR-2000; 2000US-00540217.
XX PR 23-AUG-2000; 2000US-00649167.
XX PA (HYSE-) HYSEQ INC.
XX PI Drmanac RT, Liu C, Tang YT;
XX WPI; 2001-639362/73.
XX DR N-PSDB; AAS91174.
XX New isolated polynucleotide and encoded polypeptides, useful in
PT responsible for genetic disorders or other traits and to assess
```

```
diagnostics, forensics, gene mapping, identification of mutations
PT responsible for genetic disorders or other traits and to assess
PT biodiversity.
XX Claim 20; SEQ ID NO 57346; 103pp; English.
XX The invention relates to isolated polynucleotide (I) and polypeptide (II)
CC sequences. (I) is useful as hybridisation probes, polymerase chain
CC reaction (PCR) primers, oligomers and for chromosome and gene mapping,
CC and in recombinant production of (II). The polynucleotides are also used
CC in diagnostics as expressed sequence tags for identifying expressed
CC genes. (I) is useful in gene therapy techniques to restore normal
CC activity of (II) or to treat disease states involving (II). (II) is
CC polypeptide in tissue, as molecular weight markers and as a food
CC supplement. (II) and its binding partners are useful in medical imaging
CC of sites expressing (II). (I) and (II) are useful for treating disorders
CC involving aberrant protein expression or biological activity. The
CC polypeptide and polynucleotide sequences have applications in
CC diagnostics, forensics, gene mapping, identification of mutations
CC responsible for genetic disorders or other traits to assess biodiversity
CC and to produce other types of data and products dependent on DNA and
CC amino acid sequences. ABG00010-ABG30377 represent novel human diagnostic
CC amino acid sequences. Note: The sequence data for this
CC patent did not appear in the printed specification, but was obtained in
CC electronic format directly from WIPO at
CC ftp.wipo.int/pub/published_pct_sequences
XX
XX SQ Sequence 75 AA;
Query Match 1.3%; Score 7; DB 4; Length 75;
Best Local Similarity 100.0%; Pred. No. 2.5e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 477 QGSLTDS 483
Db 66 QGSLTDS 72
|||||
RESULT 100
ABG28171
ID ABG28171 standard; protein; 75 AA.
XX AC ABG28171;
XX DT 18-FEB-2002 (first entry)
XX DE Novel human diagnostic protein #28162.
XX KW Human; chromosome mapping; gene mapping; gene therapy; forensic;
XX KW food supplement; medical imaging; diagnostic; genetic disorder.
XX OS Homo sapiens.
XX PN WO200175067-A2.
XX PD 11-OCT-2001.
XX PF 30-MAR-2001; 2001WO-US008631.
XX PR 31-MAR-2000; 2000US-00540217.
XX PR 23-AUG-2000; 2000US-00649167.
XX PA (HYSE-) HYSEQ INC.
XX PI Drmanac RT, Liu C, Tang YT;
XX WPI; 2001-639362/73.
XX DR N-PSDB; AAS92358.
XX New isolated polynucleotide and encoded polypeptides, useful in
PT diagnostics, forensics, gene mapping, identification of mutations
PT responsible for genetic disorders or other traits and to assess
```

PT biodiversity.
 XX
 PS Claim 20; SEQ ID NO 58530; 103pp; English.
 XX
 CC The invention relates to isolated polynucleotide (I) and polypeptide (II)
 CC sequences. (I) is useful as hybridisation probes, polymerase chain
 CC reaction (PCR) primers, oligomers, and for chromosome and gene mapping,
 CC and in recombinant production of (II). The polynucleotides are also used
 CC in diagnostics as expressed sequence tags for identifying expressed
 CC genes. (I) is useful in gene therapy techniques to restore normal
 CC activity of (II) or to treat disease states involving (II). (II) is
 CC useful for generating antibodies against it, detecting or quantitating a
 CC polypeptide in tissue, as molecular weight markers and as a food
 CC supplement. (II) and its binding partners are useful in medical imaging
 CC of sites expressing (II). (I) and (II) are useful for treating disorders
 CC involving aberrant protein expression or biological activity. The
 CC polypeptide and polynucleotide sequences have applications in
 CC diagnostics, forensics, gene mapping, identification of mutations
 CC responsible for genetic disorders or other traits to assess biodiversity
 CC and to produce other types of data and products dependent on DNA and
 CC amino acid sequences. ABG0010-ABG0317 represent novel human diagnostic
 CC amino acid sequences of the invention. Note: The sequence data for this
 CC patent did not appear in the printed specification, but was obtained in
 CC electronic format directly from WIPO at
 CC ftp.wipo.int/pub/published_pct_sequences
 XX
 SQ Sequence 75 AA;

Query Match 1.3%; Score 7; DB 4; length 75;
 Best Local Similarity 100.0%; Pred.No. 2.5e+02;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 477 QGSLTDS 483
 |||||
 Db 58 QGSLTDS 64

Search completed: June 7, 2004, 14:45:20
 Job time : 59 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: June 7, 2004, 14:44:26 ; Search time 20 seconds
(without alignments)
1368.088 Million cell updates/sec

Title: US-10-023-529-8

Perfect score: 530

Sequence: 1 KSSPCQPEAGPEGAQERPSQ.....APSTEASGQTGPQEPSARA 530

Scoring table: OLIGO

Gapop 60.0 , Gapext 60.0

Searched: 389414 seqs, 51625971 residues

Word size : 0

Total number of hits satisfying chosen parameters: 389414

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 300 summaries

Database : Issued Patents AA:*

- 1: /cgn2_6/ptodata/2/iaa/5A_COMB.pep.*
- 2: /cgn2_6/ptodata/2/iaa/5B_COMB.pep.*
- 3: /cgn2_6/ptodata/2/iaa/6A_COMB.pep.*
- 4: /cgn2_6/ptodata/2/iaa/6B_COMB.pep.*
- 5: /cgn2_6/ptodata/2/iaa/PCITUS_COMB.pep.*
- 6: /cgn2_6/ptodata/2/iaa/backfiles1.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	530	100.0	530	4	US-08-979-608A-8
2	530	100.0	530	4	US-09-517-849-8
3	530	100.0	530	4	US-09-616-289-8
4	429	80.9	546	4	US-09-616-289-44
5	172	32.5	557	4	US-08-979-608A-5
6	172	32.5	557	4	US-09-517-849-5
7	172	32.5	557	4	US-09-616-289-5
8	19	3.6	386	2	US-08-968-751-6
9	8	1.5	165	4	US-08-879-561-7
10	8	1.5	260	2	US-08-979-608A-41
11	7	1.3	7	4	US-09-517-849-41
12	7	1.3	7	4	US-09-616-289-41
13	7	1.3	7	4	US-09-446-350-25
14	7	1.3	77	3	US-09-446-350-25
15	7	1.3	105	4	US-09-732-210-434
16	7	1.3	124	4	US-09-489-039A-10493
17	7	1.3	127	4	US-08-849-303-19
18	7	1.3	142	4	US-09-489-039A-9418
19	7	1.3	159	3	US-09-146-950-20
20	7	1.3	172	4	US-09-079-030-117
21	7	1.3	187	4	US-09-328-352-7281
22	7	1.3	197	3	US-09-146-950-18
23	7	1.3	211	4	US-09-134-001C-5175
24	7	1.3	261	2	US-08-879-561-1
25	7	1.3	267	4	US-09-489-039A-13457
26	7	1.3	274	1	US-08-287-959-3
27	7	1.3	277	4	US-09-857-498A-7

28	7	1.3	281	4	US-09-134-000C-3980	Sequence 3980, Ap
29	7	1.3	294	4	US-09-489-039A-8327	Sequence 8327, Ap
30	7	1.3	302	4	US-09-328-352-4508	Sequence 4508, Ap
31	7	1.3	311	4	US-09-252-991A-28792	Sequence 28792, A
32	7	1.3	312	1	US-08-285-440-4	Sequence 4, Appli
33	7	1.3	312	1	US-08-630-349-4	Sequence 4, Appli
34	7	1.3	314	4	US-09-252-991A-18599	Sequence 18599, A
35	7	1.3	330	4	US-09-252-991A-28531	Sequence 28531, A
36	7	1.3	341	3	US-09-008-465-1	Sequence 1, Appli
37	7	1.3	341	4	US-09-528-959-1	Sequence 1, Appli
38	7	1.3	346	4	US-09-252-991A-27167	Sequence 27167, A
39	7	1.3	364	4	US-09-252-991A-27209	Sequence 27209, A
40	7	1.3	388	4	US-09-252-991A-21572	Sequence 21572, A
41	7	1.3	389	4	US-09-543-681A-7318	Sequence 7318, Ap
42	7	1.3	411	4	US-09-252-991A-24632	Sequence 24632, A
43	7	1.3	422	4	US-09-252-991A-21571	Sequence 21571, A
44	7	1.3	446	4	US-09-252-991A-25318	Sequence 25318, A
45	7	1.3	473	4	US-09-252-991A-16904	Sequence 16904, A
46	7	1.3	474	4	US-09-543-681A-6752	Sequence 6752, Ap
47	7	1.3	495	3	US-08-962-859A-2	Sequence 2, Appli
48	7	1.3	526	4	US-09-252-991A-20533	Sequence 20533, A
49	7	1.3	532	1	US-08-285-440-5	Sequence 5, Appli
50	7	1.3	532	1	US-08-630-349-5	Sequence 5, Appli
51	7	1.3	557	4	US-09-138-277C-3	Sequence 3, Appli
52	7	1.3	558	1	US-08-285-440-6	Sequence 6, Appli
53	7	1.3	581	1	US-08-630-349-6	Sequence 6, Appli
54	7	1.3	581	4	US-09-499-522-18	Sequence 18, Appli
55	7	1.3	582	4	US-09-269-939A-12	Sequence 12, Appli
56	7	1.3	582	4	US-09-252-991A-27626	Sequence 27626, A
57	7	1.3	630	4	US-09-499-522-16	Sequence 16, Appli
58	7	1.3	630	4	US-09-269-939A-10	Sequence 10, Appli
59	7	1.3	649	4	US-09-499-522-14	Sequence 14, Appli
60	7	1.3	649	4	US-09-269-939A-8	Sequence 8, Appli
61	7	1.3	712	4	US-09-543-681A-5392	Sequence 5392, Ap
62	7	1.3	1024	4	US-09-562-737-47	Sequence 47, Appli
63	7	1.3	1075	4	US-09-252-991A-18387	Sequence 18387, A
64	7	1.3	1141	1	US-08-131-365B-54	Sequence 54, Appli
65	7	1.3	1141	2	US-08-668-123-54	Sequence 54, Appli
66	7	1.3	1530	4	US-09-976-594-736	Sequence 736, App
67	7	1.3	1857	1	US-08-287-959-1	Sequence 1, Appli
68	7	1.3	3340	4	US-09-252-991A-23568	Sequence 23568, A
69	6	1.1	6	2	US-08-310-912A-138	Sequence 138, App
70	6	1.1	6	3	US-09-301-085-138	Sequence 138, App
71	6	1.1	6	5	PCT-US95-04589-138	Sequence 138, App
72	6	1.1	7	4	US-09-989-789-1988	Sequence 1988, Ap
73	6	1.1	9	1	US-08-339-152A-1	Sequence 1, Appli
74	6	1.1	9	2	US-08-007-999B-1	Sequence 1, Appli
75	6	1.1	9	2	US-08-689-276A-1	Sequence 1, Appli
76	6	1.1	10	3	US-08-159-339A-830	Sequence 830, App
77	6	1.1	11	3	US-08-836-075A-109	Sequence 109, App
78	6	1.1	12	3	US-08-747-599A-16	Sequence 16, Appli
79	6	1.1	15	4	US-08-979-608A-29	Sequence 29, Appli
80	6	1.1	15	4	US-09-517-849-29	Sequence 29, Appli
81	6	1.1	15	4	US-09-616-289-29	Sequence 29, Appli
82	6	1.1	16	1	US-08-408-120-15	Sequence 15, Appli
83	6	1.1	17	4	US-09-459-958-2	Sequence 2, Appli
84	6	1.1	18	6	5304631-6	Patent No. 5304631
85	6	1.1	19	2	US-08-802-322-10	Sequence 10, Appli
86	6	1.1	19	4	US-09-832-161-10	Sequence 10, Appli
87	6	1.1	20	4	US-09-834-759-544	Sequence 544, App
88	6	1.1	21	4	US-09-851-789-543	Sequence 543, App
89	6	1.1	21	4	US-08-851-843A-113	Sequence 113, App
90	6	1.1	23	3	US-08-974-549A-233	Sequence 233, App
91	6	1.1	23	3	US-08-854-050-113	Sequence 113, App
92	6	1.1	23	4	US-09-430-323-113	Sequence 113, App
93	6	1.1	23	4	US-08-912-951-114	Sequence 114, App
94	6	1.1	23	4	US-09-402-181B-233	Sequence 233, App
95	6	1.1	23	4	US-09-721-456-233	Sequence 233, App
96	6	1.1	25	2	US-08-507-598-4	Sequence 4, Appli
97	6	1.1	25	2	US-08-507-750-4	Sequence 4, Appli
98	6	1.1	25	3	US-08-764-532A-4	Sequence 4, Appli
99	6	1.1	25	3	US-08-764-538-4	Sequence 4, Appli
100	6	1.1	25	3	US-08-872-859-4	Sequence 4, Appli

101	6	1.1	30	3	US-08-927-219-41	Sequence 41, Appl	174	6	1.1	96	3	US-08-654-482-11	Sequence 11, Appl
102	6	1.1	32	2	US-08-708-620A-3	Sequence 3, Appl	175	6	1.1	98	2	US-08-341-843B-13	Sequence 13, Appl
103	6	1.1	36	1	US-08-441-534A-7	Sequence 7, Appl	176	6	1.1	98	2	US-08-427-497E-18	Sequence 18, Appl
104	6	1.1	36	1	US-08-441-534A-8	Sequence 8, Appl	177	6	1.1	100	1	US-08-473-981A-11	Sequence 11, Appl
105	6	1.1	36	1	US-08-629-752-7	Sequence 7, Appl	178	6	1.1	100	2	US-08-474-087-11	Sequence 11, Appl
106	6	1.1	36	1	US-08-629-752-8	Sequence 8, Appl	179	6	1.1	100	3	US-09-034-916-13	Sequence 13, Appl
107	6	1.1	36	2	US-08-802-991-7	Sequence 7, Appl	180	6	1.1	105	2	US-08-690-011A-15	Sequence 15, Appl
108	6	1.1	36	2	US-08-802-991-8	Sequence 8, Appl	181	6	1.1	105	4	US-09-299-495F-15	Sequence 15, Appl
109	6	1.1	36	4	US-09-079-030-61	Sequence 61, Appl	182	6	1.1	105	4	US-09-107-532A-5344	Sequence 5344, Ap
110	6	1.1	36	4	US-08-488-161-14	Sequence 14, Appl	183	6	1.1	106	4	US-09-069-023-10	Sequence 10, Appl
111	6	1.1	38	2	US-08-726-306A-101	Sequence 101, App	184	6	1.1	106	4	US-09-621-976-4439	Sequence 4439, Ap
112	6	1.1	38	3	US-09-273-685-14	Sequence 14, Appl	185	6	1.1	107	4	US-09-755-298A-2	Sequence 2, Appl
113	6	1.1	38	5	PCT-US95-11934-14	Sequence 14, Appl	186	6	1.1	109	4	US-09-574-141A-72	Sequence 72, Appl
114	6	1.1	40	1	US-08-129-129-6	Sequence 6, Appl	187	6	1.1	109	4	US-09-325-932A-48	Sequence 48, Appl
115	6	1.1	41	1	US-08-453-924-8	Sequence 8, Appl	188	6	1.1	110	4	US-09-134-000C-6399	Sequence 6399, Ap
116	6	1.1	44	1	US-08-208-008C-11	Sequence 11, Appl	189	6	1.1	112	4	US-09-107-532A-4572	Sequence 4572, Ap
117	6	1.1	45	1	US-08-339-152A-25	Sequence 25, Appl	190	6	1.1	114	2	US-08-822-262-3	Sequence 3, Appl
118	6	1.1	45	2	US-08-007-999B-8	Sequence 8, Appl	191	6	1.1	114	4	US-08-679-493A-187	Sequence 187, App
119	6	1.1	46	2	US-08-689-276A-8	Sequence 8, Appl	192	6	1.1	114	4	US-09-252-991A-18514	Sequence 18514, A
120	6	1.1	51	4	US-09-205-258-583	Sequence 583, App	193	6	1.1	115	4	US-09-376-330-22	Sequence 22, Appl
121	6	1.1	51	4	US-09-621-976-7200	Sequence 7200, Ap	194	6	1.1	117	4	US-09-134-000C-5275	Sequence 5275, Ap
122	6	1.1	52	2	US-08-164-292B-33	Sequence 33, Appl	195	6	1.1	117	4	US-09-134-000C-5589	Sequence 5589, Ap
123	6	1.1	53	3	US-08-845-623-33	Sequence 33, Appl	196	6	1.1	118	4	US-09-252-991A-23679	Sequence 23679, A
124	6	1.1	57	3	US-08-815-927-33	Sequence 33, Appl	197	6	1.1	122	3	US-08-956-307B-19	Sequence 19, Appl
125	6	1.1	57	4	US-09-103-330-33	Sequence 33, Appl	198	6	1.1	124	4	US-09-134-000C-6744	Sequence 6744, Ap
126	6	1.1	57	4	US-08-435-242-33	Sequence 33, Appl	199	6	1.1	125	4	US-09-621-976-4822	Sequence 4822, Ap
127	6	1.1	60	4	US-09-621-976-5357	Sequence 5357, Ap	200	6	1.1	128	2	US-08-928-926A-3	Sequence 3, Appl
128	6	1.1	65	4	US-09-328-352-7594	Sequence 7594, Ap	201	6	1.1	128	3	US-09-212-149-3	Sequence 3, Appl
129	6	1.1	66	3	US-09-370-473-8	Sequence 8, Appl	202	6	1.1	129	4	US-09-732-210-1307	Sequence 1307, Ap
130	6	1.1	67	4	US-09-621-976-7214	Sequence 7214, Ap	203	6	1.1	130	4	US-09-732-210-1308	Sequence 1308, Ap
131	6	1.1	67	4	US-09-621-976-7215	Sequence 7215, Ap	204	6	1.1	132	4	US-09-716-129-57	Sequence 57, Appl
132	6	1.1	68	4	US-09-621-976-5028	Sequence 5028, Ap	205	6	1.1	133	4	US-09-732-210-82	Sequence 82, Appl
133	6	1.1	73	4	US-08-580-043B-2	Sequence 2, Appl	206	6	1.1	133	4	US-09-489-039A-13255	Sequence 13255, A
134	6	1.1	74	3	US-08-851-843A-187	Sequence 187, App	207	6	1.1	134	4	US-09-252-991A-24342	Sequence 24342, A
135	6	1.1	74	3	US-08-974-549A-306	Sequence 306, App	208	6	1.1	135	4	US-09-328-352-5695	Sequence 5695, Ap
136	6	1.1	74	3	US-08-836-075A-10	Sequence 10, Appl	209	6	1.1	137	4	US-09-732-210-122	Sequence 122, App
137	6	1.1	74	3	US-08-854-050-187	Sequence 187, App	210	6	1.1	137	6	5202428-9	Patent No. 5202428
138	6	1.1	74	4	US-09-430-323-187	Sequence 187, App	211	6	1.1	138	3	US-08-836-075A-60	Sequence 60, Appl
139	6	1.1	74	4	US-08-635-886C-198	Sequence 198, App	212	6	1.1	140	4	US-09-540-236-2611	Sequence 2611, Ap
140	6	1.1	74	4	US-09-402-161B-306	Sequence 306, App	213	6	1.1	141	3	US-09-230-944-4	Sequence 4, Appl
141	6	1.1	74	4	US-08-974-690C-198	Sequence 198, App	214	6	1.1	141	4	US-09-873-233A-4	Sequence 4, Appl
142	6	1.1	74	4	US-09-721-456-306	Sequence 306, App	215	6	1.1	142	4	US-09-543-681A-4478	Sequence 4478, Ap
143	6	1.1	75	1	US-07-934-017-1	Sequence 1, Appl	216	6	1.1	142	4	US-09-543-681A-6036	Sequence 6036, Ap
144	6	1.1	75	4	US-09-252-991A-29599	Sequence 29599, A	217	6	1.1	142	4	US-09-489-039A-8382	Sequence 8382, Ap
145	6	1.1	75	4	US-09-732-210-833	Sequence 833, App	218	6	1.1	142	4	US-09-489-039A-11181	Sequence 11181, A
146	6	1.1	76	1	US-08-168-941-1	Sequence 1, Appl	219	6	1.1	143	4	US-09-387-418A-17	Sequence 17, Appl
147	6	1.1	79	4	US-09-134-000C-5581	Sequence 5581, Ap	220	6	1.1	144	4	US-09-732-210-1406	Sequence 1406, Ap
148	6	1.1	80	4	US-09-107-532A-5588	Sequence 5588, Ap	221	6	1.1	144	4	US-09-489-039A-11477	Sequence 11477, A
149	6	1.1	82	2	US-08-690-011A-13	Sequence 13, Appl	222	6	1.1	145	4	US-09-489-039A-10614	Sequence 10614, A
150	6	1.1	82	3	US-09-905-223-398	Sequence 398, App	223	6	1.1	146	2	US-07-885-0893-37	Sequence 37, Appl
151	6	1.1	82	4	US-09-299-495F-13	Sequence 13, Appl	224	6	1.1	146	4	US-09-252-991A-21410	Sequence 21410, A
152	6	1.1	83	4	US-09-732-210-1584	Sequence 1584, Ap	225	6	1.1	147	4	US-09-252-991A-21269	Sequence 21269, A
153	6	1.1	83	4	US-09-732-210-1587	Sequence 1587, Ap	226	6	1.1	148	4	US-09-621-976-4708	Sequence 4708, Ap
154	6	1.1	83	4	US-09-134-000C-5875	Sequence 5875, Ap	227	6	1.1	149	4	US-09-489-039A-10664	Sequence 10664, A
155	6	1.1	83	4	US-09-621-976-7069	Sequence 7069, Ap	228	6	1.1	151	4	US-09-489-039A-13318	Sequence 13318, A
156	6	1.1	84	1	US-08-510-039-2	Sequence 2, Appl	229	6	1.1	152	1	US-08-046-583-11	Sequence 11, Appl
157	6	1.1	84	1	US-07-748-510-2	Sequence 2, Appl	230	6	1.1	152	1	US-08-261-677-11	Sequence 11, Appl
158	6	1.1	88	4	US-09-134-000C-5567	Sequence 5567, Ap	231	6	1.1	152	1	US-08-384-556A-2	Sequence 2, Appl
159	6	1.1	89	4	US-08-618-485B-2	Sequence 2, Appl	232	6	1.1	152	1	US-08-384-556A-7	Sequence 7, Appl
160	6	1.1	90	4	US-09-107-532A-6712	Sequence 6712, Ap	233	6	1.1	152	1	US-08-592-936B-17	Sequence 17, Appl
161	6	1.1	91	4	US-09-732-210-1218	Sequence 1218, Ap	234	6	1.1	152	2	US-07-885-0893-36	Sequence 36, Appl
162	6	1.1	92	4	US-09-489-039A-11823	Sequence 11823, A	235	6	1.1	152	2	US-08-331-355A-23	Sequence 23, Appl
163	6	1.1	93	4	US-08-936-165A-381	Sequence 381, App	236	6	1.1	152	2	US-08-111-573-17	Sequence 17, Appl
164	6	1.1	94	2	US-08-737-716-10	Sequence 10, App	237	6	1.1	152	2	US-09-071-035-128	Sequence 128, App
165	6	1.1	94	4	US-09-099-041A-31	Sequence 31, Appl	238	6	1.1	152	5	PCT-US94-12364-11	Sequence 11, Appl
166	6	1.1	94	4	US-09-245-281-31	Sequence 31, Appl	239	6	1.1	152	5	PCT-US94-12364-23	Sequence 23, Appl
167	6	1.1	94	4	US-08-618-485B-3	Sequence 3, Appl	240	6	1.1	152	5	PCT-US95-07753-2	Sequence 2, Appl
168	6	1.1	94	4	US-09-207-359B-31	Sequence 31, Appl	241	6	1.1	152	5	PCT-US95-07753-7	Sequence 7, Appl
169	6	1.1	94	4	US-09-340-620A-31	Sequence 31, Appl	242	6	1.1	153	4	US-09-199-637A-213	Sequence 213, App
170	6	1.1	94	4	US-09-252-991A-26448	Sequence 26448, A	243	6	1.1	154	4	US-09-252-991A-20709	Sequence 20709, A
171	6	1.1	94	4	US-09-865-364-31	Sequence 31, Appl	244	6	1.1	154	4	US-09-107-532A-5411	Sequence 5411, Ap
172	6	1.1	94	4	US-09-673-395A-415	Sequence 415, App	245	6	1.1	155	1	US-08-328-961-8	Sequence 8, Appl
173	6	1.1	96	3	US-09-258-373-22	Sequence 22, Appl	246	6	1.1				

247 6 1.1 155 1 US-08-462-397-8
248 6 1.1 155 3 US-08-989-251-39
249 6 1.1 155 3 US-09-340-250-39
250 6 1.1 155 4 US-09-528-108-39
251 6 1.1 155 4 US-09-732-210-1237
252 6 1.1 157 4 US-09-252-991A-20847
253 6 1.1 160 4 US-09-252-991A-28043
254 6 1.1 160 4 US-09-198-452A-1244
255 6 1.1 162 4 US-09-198-452A-1190
256 6 1.1 164 4 US-09-621-976-5067
257 6 1.1 165 4 US-09-252-991A-27342
258 6 1.1 166 4 US-09-134-000C-5854
259 6 1.1 168 4 US-09-252-991A-30271
260 6 1.1 168 4 US-09-107-532A-3745
261 6 1.1 169 3 US-08-928-941D-35
262 6 1.1 169 4 US-09-280-590A-45
263 6 1.1 169 4 US-09-892-398-45
264 6 1.1 172 4 US-09-107-532A-5616
265 6 1.1 174 4 US-09-328-352-7878
266 6 1.1 175 4 US-09-543-681A-8208
267 6 1.1 175 4 US-09-543-681A-8328
268 6 1.1 176 4 US-09-387-418A-16
269 6 1.1 176 4 US-09-252-991A-21594
270 6 1.1 179 4 US-09-252-991A-24507
271 6 1.1 180 4 US-09-252-991A-18339
272 6 1.1 180 4 US-09-252-991A-23183
273 6 1.1 180 4 US-09-252-991A-26449
274 6 1.1 181 1 US-08-034-245-14
275 6 1.1 181 4 US-09-325-932A-180
276 6 1.1 181 4 US-09-252-991A-19502
277 6 1.1 181 4 US-09-134-000C-5492
278 6 1.1 182 4 US-09-198-452A-164
279 6 1.1 182 4 US-09-543-681A-5698
280 6 1.1 182 4 US-09-134-000C-6084
281 6 1.1 183 4 US-09-071-035-126
282 6 1.1 183 4 US-09-489-039A-10037
283 6 1.1 183 4 US-09-134-000C-4534
284 6 1.1 185 4 US-09-198-452A-1209
285 6 1.1 185 4 US-09-418-710-64
286 6 1.1 187 4 US-09-252-991A-29442
287 6 1.1 188 4 US-09-107-532A-7007
288 6 1.1 191 3 US-08-989-251-41
289 6 1.1 191 3 US-09-340-250-41
290 6 1.1 191 4 US-09-528-108-41
291 6 1.1 192 4 US-09-489-039A-10990
292 6 1.1 192 4 US-09-134-000C-3560
293 6 1.1 194 4 US-09-252-991A-22009
294 6 1.1 195 4 US-09-252-991A-21451
295 6 1.1 195 4 US-09-489-039A-7491
296 6 1.1 196 4 US-09-540-236-3127
297 6 1.1 199 4 US-09-252-991A-18507
298 6 1.1 202 4 US-08-311-731A-13
299 6 1.1 204 4 US-09-230-670C-8
300 6 1.1 207 4 US-09-252-991A-28701

ALIGNMENTS

RESULT 1
US-08-979-608A-8
; Sequence 8, Application US/08979608A
; Patent No. 6355451
; GENERAL INFORMATION:
; APPLICANT: Lees, Ann M.
; Lees, Robert S.
; Law, Simon W.
; Arjona, Anibal A.
; TITLE OF INVENTION: NOVEL LOW DENSITY LIPOPROTEIN
; BINDING PROTEINS AND THEIR USES IN DIAGNOSING AND
; TREATING ATHEROSCLEROSIS
; NUMBER OF SEQUENCES: 42
; CORRESPONDENCE ADDRESS:

; ; ADDRESS: Fish & Richardson P.C.
; ; STREET: 225 Franklin Street
; ; CITY: Boston
; ; STATE: MA
; ; COUNTRY: USA
; ; ZIP: 02110-2804
; ; COMPUTER READABLE FORM:
; ; MEDIUM TYPE: Diskette
; ; COMPUTER: IBM Compatible
; ; OPERATING SYSTEM: DOS
; ; SOFTWARE: FastSeq for Windows Version 2.0
; ; CURRENT APPLICATION DATA:
; ; APPLICATION NUMBER: US/08/979,608A
; ; FILING DATE: 26-Nov. 6355451-1997
; ; PRIOR APPLICATION DATA:
; ; APPLICATION NUMBER: US 60/048,547
; ; FILING DATE: 03-JUN-1997
; ; APPLICATION NUMBER: US 60/031,930
; ; FILING DATE: 27-NOV-1996
; ; ATTORNEY/AGENT INFORMATION:
; ; NAME: Myers, Louis
; ; REGISTRATION NUMBER: 35,965
; ; REFERENCE/DOCKET NUMBER: 10797-002001 (formerly 3983/59818)
; ; TELECOMMUNICATION INFORMATION:
; ; TELEPHONE: 617/542-5070
; ; TELEFAX: 617/542-8906
; ; INFORMATION FOR SEQ ID NO: 8:
; ; SEQUENCE CHARACTERISTICS:
; ; LENGTH: 530 amino acids
; ; TYPE: amino acid
; ; TOPOLOGY: linear
; ; MOLECULE TYPE: protein
; ; SEQUENCE DESCRIPTION: SEQ ID NO: 8:
; ; US-08-979-608A-8
; ; Query Match 100.0%; Score 530; DB 4; Length 530;
; ; Best Local Similarity 100.0%; Pred. No. 0;
; ; Matches 530; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
; ; QY 1 KSPGQPEAGPEGAQPSQAAPAVEAGPGSSQAPKPEGAQARTAGSGALRVSELS 60
; ; DB 1 KSPGQPEAGPEGAQPSQAAPAVEAGPGSSQAPKPEGAQARTAGSGALRVSELS 60
; ; QY 61 RQLEDIILSTYCVNNOGPGEDGAGPEAPDEAKSRTYVARNGEPEPTPVYGEKPS 120
; ; DB 61 RQLEDIILSTYCVNNOGPGEDGAGPEAPDEAKSRTYVARNGEPEPTPVYGEKPS 120
; ; QY 121 KGPNTBEIRQSDVGRDRHRRPOEKKAKGLGKEITLLMOTLNTLSTPEKLAALCKKY 180
; ; DB 121 KGPNTBEIRQSDVGRDRHRRPOEKKAKGLGKEITLLMOTLNTLSTPEKLAALCKKY 180
; ; QY 181 AELLEHHNSOKMKLLQKQSQVLQVQKHRLRGHSHKAVLARSKLSLCRELQHNRLK 240
; ; DB 181 AELLEHHNSOKMKLLQKQSQVLQVQKHRLRGHSHKAVLARSKLSLCRELQHNRLK 240
; ; QY 241 EGVQRAPEEERKEVTSHPQVTLNDIQLQMEQHNRNSKLRQENMELARLKKLIQY 300
; ; DB 241 EGVQRAPEEERKEVTSHPQVTLNDIQLQMEQHNRNSKLRQENMELARLKKLIQY 300
; ; QY 301 ELRREHIDKVFHKDLOOQLVDKQLQQAQEMLEKAEERHOREKDFLLKEAVESQRMCELM 360
; ; DB 301 ELRREHIDKVFHKDLOOQLVDKQLQQAQEMLEKAEERHOREKDFLLKEAVESQRMCELM 360
; ; QY 361 KQOETHLKKQALALYTEKPEEPQNTLSKSSSEVFTTFKQEMENKTKIKKLEKETTMYRSRW 420
; ; DB 361 KQOETHLKKQALALYTEKPEEPQNTLSKSSSEVFTTFKQEMENKTKIKKLEKETTMYRSRW 420
; ; QY 421 ESSNKALLEAEKBTVRDKELEGQVKIQRLEKLCRALQTERNDLNKRVQDLSAGGQSL 480
; ; DB 421 ESSNKALLEAEKBTVRDKELEGQVKIQRLEKLCRALQTERNDLNKRVQDLSAGGQSL 480
; ; QY 481 TDSGPRRPEGPGACPSPPVTEAPCVPGAPSTEASGQTPGPTSARA 530

Db 481 TDSGPERPEGGAQAPSSPRVTEAPCYPGAPSTEASGQTGPQPTSA 530

RESULT 2

US-09-517-849-8
; Sequence 8, Application US/09517849
; Patent No. 5605588
; GENERAL INFORMATION:
; APPLICANT: Lees, Ann M.
; Lees, Robert S.
; Law, Simon W.
; Arjona, Anibal A.
; TITLE OF INVENTION: NOVEL LOW DENSITY LIPOPROTEIN
; BINDING PROTEINS AND THEIR USES IN DIAGNOSING AND
; TREATING ATHEROSCLEROSIS
; NUMBER OF SEQUENCES: 42
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Fish & Richardson P.C.
; STREET: 225 Franklin Street
; CITY: Boston
; STATE: MA
; COUNTRY: USA
; ZIP: 02110-2804
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/517,849
; FILING DATE: 02-Mar-2000
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/379,608
; FILING DATE: 26-NOV-1997
; ATTORNEY/AGENT INFORMATION:
; NAME: Myers, Louis
; REGISTRATION NUMBER: 35,965
; REFERENCE/DOCKET NUMBER: 10797-003001
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 617/542-5070
; TELEFAX: 617/542-8906
; INFORMATION FOR SEQ ID NO: 8:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 530 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; SEQUENCE DESCRIPTION: SEQ ID NO: 8:
US-09-517-849-8

Query Match 100.0%; Score 530; DB 4; Length 530;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 530; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 1 KSSPGPEAGPEGAQRPSSQAAPAVEAGPGSSQAAPKPEGAQARTAQSGALRDVSELS 60
Db 1 KSSPGPEAGPEGAQRPSSQAAPAVEAGPGSSQAAPKPEGAQARTAQSGALRDVSELS 60
Qy 61 RQLEDILSTYCVDDNQGPGEDGAQGEPAEPEDAESKRTTYVARNGEPEPTPVVYGEKPS 120
Db 61 RQLEDILSTYCVDDNQGPGEDGAQGEPAEPEDAESKRTTYVARNGEPEPTPVVYGEKPS 120
Qy 121 KGPNTTEIRQSDVGDHRRPQPKKAKGLGKEITLLMTLTNTLSTPBKLAALCKKY 180
Db 121 KGPNTTEIRQSDVGDHRRPQPKKAKGLGKEITLLMTLTNTLSTPBKLAALCKKY 180
Qy 181 AELLEHRNSQOMKLLQKQSOLVQEKDHLRGHSAVLARSKLSLCRELQHRNRSJK 240
Db 181 AELLEHRNSQOMKLLQKQSOLVQEKDHLRGHSAVLARSKLSLCRELQHRNRSJK 240
Qy 241 EGVQARAREEERKEVTSHFQVTLNDIQLQMEQHNRNSKLROENMELAEERLKKLIEQY 300
Db 241 EGVQARAREEERKEVTSHFQVTLNDIQLQMEQHNRNSKLROENMELAEERLKKLIEQY 300

Qy 301 ELREEHIDKVPKHDLQQOLVDKLCQAQEMLKBAERHOREKDFLLKEAVSORMCELM 360
Db 301 ELREEHIDKVPKHDLQQOLVDKLCQAQEMLKBAERHOREKDFLLKEAVSORMCELM 360
Qy 361 KQETHLKOALALYTEKFEFPQNTLSKSEVPTTTPKQEMEKMTKKIKLEKETTMYRSW 420
Db 361 KQETHLKOALALYTEKFEFPQNTLSKSEVPTTTPKQEMEKMTKKIKLEKETTMYRSW 420
Qy 421 ESSNKALLEWAEKTVRDKELEGLOVKIORLKLALOTERNDLNKRVDLSAGQGSGL 480
Db 421 ESSNKALLEWAEKTVRDKELEGLOVKIORLKLALOTERNDLNKRVDLSAGQGSGL 480
Qy 481 TDSGPERPEGGAQAPSSPRVTEAPCYPGAPSTEASGQTGPQPTSA 530
Db 481 TDSGPERPEGGAQAPSSPRVTEAPCYPGAPSTEASGQTGPQPTSA 530

RESULT 3

US-09-616-289-8
; Sequence 8, Application US/09616289
; Patent No. 5632923
; GENERAL INFORMATION:
; APPLICANT: Lees, Ann M.
; APPLICANT: Lees, Robert S.
; APPLICANT: Law, Simon W.
; APPLICANT: Arjona, Anibal A.
; TITLE OF INVENTION: NOVEL LOW DENSITY LIPOPROTEIN BINDING
; TITLE OF INVENTION: PROTEINS AND THEIR USE IN DIAGNOSING AND TREATING
; TITLE OF INVENTION: ATHEROSCLEROSIS
; FILE REFERENCE: 10797-004001
; CURRENT APPLICATION NUMBER: US/09/616,289
; CURRENT FILING DATE: 2000-07-14
; PRIOR APPLICATION NUMBER: US 09/517,849
; PRIOR FILING DATE: 2000-03-02
; PRIOR APPLICATION NUMBER: US 08/379,608
; PRIOR FILING DATE: 1997-11-26
; PRIOR APPLICATION NUMBER: US 60/031,930
; PRIOR FILING DATE: 1996-11-27
; PRIOR APPLICATION NUMBER: US 60/048,547
; PRIOR FILING DATE: 1997-06-03
; NUMBER OF SEQ ID NOS: 53
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 8
; LENGTH: 530
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-616-289-8

Query Match 100.0%; Score 530; DB 4; Length 530;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 530; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 1 KSSPGPEAGPEGAQRPSSQAAPAVEAGPGSSQAAPKPEGAQARTAQSGALRDVSELS 60
Db 1 KSSPGPEAGPEGAQRPSSQAAPAVEAGPGSSQAAPKPEGAQARTAQSGALRDVSELS 60
Qy 61 RQLEDILSTYCVDDNQGPGEDGAQGEPAEPEDAESKRTTYVARNGEPEPTPVVYGEKPS 120
Db 61 RQLEDILSTYCVDDNQGPGEDGAQGEPAEPEDAESKRTTYVARNGEPEPTPVVYGEKPS 120
Qy 121 KGPNTTEIRQSDVGDHRRPQPKKAKGLGKEITLLMTLTNTLSTPBKLAALCKKY 180
Db 121 KGPNTTEIRQSDVGDHRRPQPKKAKGLGKEITLLMTLTNTLSTPBKLAALCKKY 180
Qy 181 AELLEHRNSQOMKLLQKQSOLVQEKDHLRGHSAVLARSKLSLCRELQHRNRSJK 240
Db 181 AELLEHRNSQOMKLLQKQSOLVQEKDHLRGHSAVLARSKLSLCRELQHRNRSJK 240
Qy 241 EGVQARAREEERKEVTSHFQVTLNDIQLQMEQHNRNSKLROENMELAEERLKKLIEQY 300
Db 241 EGVQARAREEERKEVTSHFQVTLNDIQLQMEQHNRNSKLROENMELAEERLKKLIEQY 300

QY 301 ELREEHIDKVFHKDLQOQVDAKLQQAQEMLKEAERHOREKDFLLKEAVESQRMCELM 360
DB 301 ELREEHIDKVFHKDLQOQVDAKLQQAQEMLKEAERHOREKDFLLKEAVESQRMCELM 360
QY 361 KQOETHLKOALALYTEKPEFQNTLSKSSSEVFTTFKQEMEMKTKKIKLEKETTMYRSRW 420
DB 361 KQOETHLKOALALYTEKPEFQNTLSKSSSEVFTTFKQEMEMKTKKIKLEKETTMYRSRW 420
QY 421 ESSNKALLEMAEKTVDKLEGLQVQIKORLEKLCRALQTERNDLNKRVQDLSAGGQSL 480
DB 421 ESSNKALLEMAEKTVDKLEGLQVQIKORLEKLCRALQTERNDLNKRVQDLSAGGQSL 480
QY 481 TDSGPERRPEGCAQAPSSPRVTEAPCYGAPSTASGQTGPQPTTSARA 530
DB 481 TDSGPERRPEGCAQAPSSPRVTEAPCYGAPSTASGQTGPQPTTSARA 530

RESULT 4

US-09-616-289-44
; Sequence 44, Application US/09616289
; Patent No. 6632923
; GENERAL INFORMATION:
; APPLICANT: Lees, Ann M.
; APPLICANT: Lees, Robert S.
; APPLICANT: Law, Simon W.
; APPLICANT: Arjona, Anibal A.
; TITLE OF INVENTION: NOVEL LOW DENSITY LIPOPROTEIN BINDING
; TITLE OF INVENTION: PROTEINS AND THEIR USE IN DIAGNOSING AND TREATING
; FILE REFERENCE: 10797-004001
; CURRENT APPLICATION NUMBER: US/09/616,289
; CURRENT FILING DATE: 2000-07-14
; PRIOR APPLICATION NUMBER: US 09/517,849
; PRIOR FILING DATE: 2000-03-02
; PRIOR APPLICATION NUMBER: US 08/979,608
; PRIOR FILING DATE: 1997-11-26
; PRIOR APPLICATION NUMBER: US 60/031,930
; PRIOR FILING DATE: 1996-11-27
; PRIOR APPLICATION NUMBER: US 60/048,547
; PRIOR FILING DATE: 1997-06-03
; NUMBER OF SEQ ID NOS: 53
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 44
; LENGTH: 546
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-616-289-44

Query Match 80.9%; Score 429; DB 4; Length 546;
Best Local Similarity 99.8%; Pred. No. 0;
Matches 529; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 KSSPGQPEAGPQAGPQAPPAVEAEGPGSSQAPRKEGQAARTAGSGLRDVSEELS 60
DB 17 KSSPGQPEAGPQAGPQAPPAVEAEGPGSSQAPRKEGQAARTAGSGLRDVSEELS 76
QY 61 RQLEDILSTYCVNNQGGEDGAGCEPAEPEDAESRTTYVARNGEPEPTPVVGEKPS 120
DB 77 RQLEDILSTYCVNNQGGEDGAGCEPAEPEDAESRTTYVARNGEPEPTPVVGEKPS 136
QY 121 KGPNTETIRQSDVGDHRPQEKKAAGLQKBITLLMOTLNTLSTPEEKLAALCKKY 180
DB 137 KGPNTETIRQSDVGDHRPQEKKAAGLQKBITLLMOTLNTLSTPEEKLAALCKKY 196
QY 181 AELLERHNSQOMKLLQKQSQOLVQEKDHLGHSKAVLARSKLESCLREIQRNRSK 240
DB 197 AELLERHNSQOMKLLQKQSQOLVQEKDHLGHSKAVLARSKLESCLREIQRNRSK 256
QY 241 EGVQARSEEBEERKREVTSHFQVTLNDIQLQMEQHNRNSKLRQENMELAEKRLIKY 300
DB 257 EGVQARSEEBEERKREVTSHFQVTLNDIQLQMEQHNRNSKLRQENMELAEKRLIKY 316
QY 301 ELREEHIDKVFHKDLQOQVDAKLQQAQEMLKEAERHOREKDFLLKEAVESQRMCELM 360

DB 317 ELREEHIDKVFHKDLQOQVDAKLQQAQEMLKEAERHOREKDFLLKEAVESQRMCELM 376
QY 361 KQOETHLKOALALYTEKPEFQNTLSKSSSEVFTTFKQEMEMKTKKIKLEKETTMYRSRW 420
DB 377 KQOETHLKOALALYTEKPEFQNTLSKSSSEVFTTFKQEMEMKTKKIKLEKETTMYRSRW 436
QY 421 ESSNKALLEMAEKTVDKLEGLQVQIKORLEKLCRALQTERNDLNKRVQDLSAGGQSL 480
DB 437 ESSNKALLEMAEKTVDKLEGLQVQIKORLEKLCRALQTERNDLNKRVQDLSAGGQSL 496
QY 481 TDSGPERRPEGCAQAPSSPRVTEAPCYGAPSTASGQTGPQPTTSARA 530
DB 497 TDSGPERRPEGCAQAPSSPRVTEAPCYGAPSTASGQTGPQPTTSARA 546

RESULT 5

US-08-979-608A-5
; Sequence 5, Application US/08979608A
; Patent No. 6355451
; GENERAL INFORMATION:
; APPLICANT: Lees, Ann M.
; APPLICANT: Lees, Robert S.
; APPLICANT: Law, Simon W.
; APPLICANT: Arjona, Anibal A.
; TITLE OF INVENTION: NOVEL LOW DENSITY LIPOPROTEIN
; BINDING PROTEINS AND THEIR USES IN DIAGNOSING AND
; TREATING ATHEROSCLEROSIS
; NUMBER OF SEQUENCES: 42
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Fish & Richardson P.C.
; STREET: 225 Franklin Street
; CITY: Boston
; STATE: MA
; COUNTRY: USA
; ZIP: 02110-2804
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/979,608A
; FILING DATE: 26-NOV-1997
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 60/048,547
; FILING DATE: 03-JUN-1997
; APPLICATION NUMBER: US 60/031,930
; FILING DATE: 27-NOV-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: Myers, Louis
; REGISTRATION NUMBER: 35,965
; REFERENCE/DOCKET NUMBER: 10797-002001 (formerly 3983/59818)
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 617/542-5070
; TELEFAX: 617/542-8906
; INFORMATION FOR SEQ ID NO: 5:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 557 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; FRAGMENT TYPE: internal
; SEQUENCE DESCRIPTION: SEQ ID NO: 5:
US-08-979-608A-5

Query Match 32.5%; Score 172; DB 4; Length 557;
Best Local Similarity 100.0%; Pred. No. 1.3e-149;
Matches 172; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 264 TLNDIQLQMEQHNRNSKLRQENMELAEKRLIKYEQYELREEHIDKVFHKDLQOQVDA 323
DB 280 TLNDIQLQMEQHNRNSKLRQENMELAEKRLIKYEQYELREEHIDKVFHKDLQOQVDA 339

Qy 324 KLOQAQEMLKEAERHOREKDFLLKEAVESQRMCELMKQOETHLKQOALYTEKFEFQ 383
Db 340 KLOQAQEMLKEAERHOREKDFLLKEAVESQRMCELMKQOETHLKQOALYTEKFEFQ 399
Qy 384 TLSKSSEVFTTFKQEMEKTKIKKLEKETTMYRSWESSNKALLEMAEKT 435
Db 400 TLSKSSEVFTTFKQEMEKTKIKKLEKETTMYRSWESSNKALLEMAEKT 451

RESULT 6
US-09-517-849-5
; Sequence 5, Application US/09517849
; Patent No. 6605588
; GENERAL INFORMATION:
; APPLICANT: Lees, Ann M.
; APPLICANT: Lees, Robert S.
; APPLICANT: Law, Simon W.
; APPLICANT: Arjona, Anibal A.
; TITLE OF INVENTION: NOVEL LOW DENSITY LIPOPROTEIN
; BINDING PROTEINS AND THEIR USES IN DIAGNOSING AND
; TREATING ATHEROSCLEROSIS
; NUMBER OF SEQUENCES: 42
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Fish & Richardson P.C.
; STREET: 225 Franklin Street
; CITY: Boston
; STATE: MA
; COUNTRY: USA
; ZIP: 02110-2804
; COMPUTER READABLE FORM:
; MEDIUM TYPE: diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/517,849
; FILING DATE: 02-Mar-2000
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/979,608
; FILING DATE: 26-NOV-1997
; ATTORNEY/AGENT INFORMATION:
; NAME: Myers, Louis
; REGISTRATION NUMBER: 35,965
; REFERENCE/DOCKET NUMBER: 10797-003001
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 617/542-5070
; TELEFAX: 617/542-8906
; INFORMATION FOR SEQ ID NO: 5:
; SEQUENCE DESCRIPTION: SEQ ID NO: 5:
; LENGTH: 557 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; FRAGMENT TYPE: internal
; SEQUENCE DESCRIPTION: SEQ ID NO: 5:
US-09-517-849-5

Query Match 32.5%; Score 172; DB 4; Length 557;
Best Local Similarity 100.0%; Pred. No. 1.3e-149;
Matches 172; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 264 TLNDIQLQMEQHNRNSKLQENMELAEKLLIEQYELREHIDKVFKHDLQOQLVDA 323
Db 280 TLNDIQLQMEQHNRNSKLQENMELAEKLLIEQYELREHIDKVFKHDLQOQLVDA 339
Qy 324 KLOQAQEMLKEAERHOREKDFLLKEAVESQRMCELMKQOETHLKQOALYTEKFEFQ 383
Db 340 KLOQAQEMLKEAERHOREKDFLLKEAVESQRMCELMKQOETHLKQOALYTEKFEFQ 399
Qy 384 TLSKSSEVFTTFKQEMEKTKIKKLEKETTMYRSWESSNKALLEMAEKT 435
Db 400 TLSKSSEVFTTFKQEMEKTKIKKLEKETTMYRSWESSNKALLEMAEKT 451

RESULT 8
US-08-968-751-6
; Sequence 6, Application US/08968751
; Patent No. 5948643
; GENERAL INFORMATION:
; APPLICANT: Rubinfeld, Bonnie
; APPLICANT: Polakis, Paul G.
; APPLICANT: Ligenfelter, Carol
; APPLICANT: Vuong, Terilyn T.
; TITLE OF INVENTION: MODULATORS OF BRCA1 ACTIVITY
; NUMBER OF SEQUENCES: 6
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: ONYX Pharmaceuticals, Inc.
; STREET: 3031 Research Drive
; CITY: Richmond
; STATE: CA
; COUNTRY: USA
; ZIP: 94806
; COMPUTER READABLE FORM: disk
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30

RESULT 7
US-09-616-289-5
; Sequence 5, Application US/09616289
; Patent No. 6632923
; GENERAL INFORMATION:
; APPLICANT: Lees, Ann M.
; APPLICANT: Lees, Robert S.
; APPLICANT: Law, Simon W.
; APPLICANT: Arjona, Anibal A.
; TITLE OF INVENTION: NOVEL LOW DENSITY LIPOPROTEIN BINDING
; PROTEINS AND THEIR USE IN DIAGNOSING AND TREATING
; ATHEROSCLEROSIS
; FILE REFERENCE: 10797-004001
; CURRENT APPLICATION NUMBER: US/09/616,289
; PRIOR FILING DATE: 2000-07-14
; PRIOR APPLICATION NUMBER: US 09/517,849
; PRIOR FILING DATE: 2000-03-02
; PRIOR APPLICATION NUMBER: US 08/979,608
; PRIOR FILING DATE: 1997-11-26
; PRIOR APPLICATION NUMBER: US 60/031,930
; PRIOR FILING DATE: 1996-11-27
; PRIOR APPLICATION NUMBER: US 60/048,547
; PRIOR FILING DATE: 1997-06-03
; NUMBER OF SEQ ID NOS: 53
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 5
; LENGTH: 557
; TYPE: PRT
; ORGANISM: Oryctolagus cuniculus
US-09-616-289-5

Query Match 32.5%; Score 172; DB 4; Length 557;
Best Local Similarity 100.0%; Pred. No. 1.3e-149;
Matches 172; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 264 TLNDIQLQMEQHNRNSKLQENMELAEKLLIEQYELREHIDKVFKHDLQOQLVDA 323
Db 280 TLNDIQLQMEQHNRNSKLQENMELAEKLLIEQYELREHIDKVFKHDLQOQLVDA 339
Qy 324 KLOQAQEMLKEAERHOREKDFLLKEAVESQRMCELMKQOETHLKQOALYTEKFEFQ 383
Db 340 KLOQAQEMLKEAERHOREKDFLLKEAVESQRMCELMKQOETHLKQOALYTEKFEFQ 399
Qy 384 TLSKSSEVFTTFKQEMEKTKIKKLEKETTMYRSWESSNKALLEMAEKT 435
Db 400 TLSKSSEVFTTFKQEMEKTKIKKLEKETTMYRSWESSNKALLEMAEKT 451

RESULT 8
US-08-968-751-6
; Sequence 6, Application US/08968751
; Patent No. 5948643
; GENERAL INFORMATION:
; APPLICANT: Rubinfeld, Bonnie
; APPLICANT: Polakis, Paul G.
; APPLICANT: Ligenfelter, Carol
; APPLICANT: Vuong, Terilyn T.
; TITLE OF INVENTION: MODULATORS OF BRCA1 ACTIVITY
; NUMBER OF SEQUENCES: 6
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: ONYX Pharmaceuticals, Inc.
; STREET: 3031 Research Drive
; CITY: Richmond
; STATE: CA
; COUNTRY: USA
; ZIP: 94806
; COMPUTER READABLE FORM: disk
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30

CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/968,751
FILING DATE:
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Giotta, Gregory
REGISTRATION NUMBER: 32,028
REFERENCE/DOCKET NUMBER: ONYX1024 GG
TELECOMMUNICATION INFORMATION:
TELEPHONE: (510) 262-8710
TELEFAX: (510) 222-9758
INFORMATION FOR SEQ ID NO: 6:
SEQUENCE CHARACTERISTICS:
LENGTH: 386 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-968-751-6

Query Match
Best Local Similarity 100.0%; Pred. No. 2.9e-09; Length 386;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 163 LNTLSTPEKLAALCKKYA 181
Db 135 LNTLSTPEKLAALCKKYA 153

RESULT 9
US-09-252-991A-19490
Sequence 19490, Application US/09252991A
Patent No. 6551795
GENERAL INFORMATION:
APPLICANT: Marc J. Rubenfield et al.
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
FILE REFERENCE: 107196.136
CURRENT APPLICATION NUMBER: US/09/252,991A
CURRENT FILING DATE: 1999-02-18
PRIOR APPLICATION NUMBER: US 60/074,788
PRIOR FILING DATE: 1998-02-18
PRIOR APPLICATION NUMBER: US 60/094,190
PRIOR FILING DATE: 1998-07-27
NUMBER OF SEQ ID NOS: 33142
SEQ ID NO 19490
LENGTH: 165
TYPE: PRT
ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-19490

Query Match
Best Local Similarity 100.0%; Pred. No. 18; Length 155;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 137 DRDHRPQ 144
Db 40 DRDHRPQ 47

RESULT 10
US-08-879-561-7
Sequence 7, Application US/08879561
Patent No. 5817482
GENERAL INFORMATION:
APPLICANT: Bandman, Olga
APPLICANT: Hillman, Jennifer L.
APPLICANT: Hawkins, Phillip R.
APPLICANT: Guegler, Karl J.
APPLICANT: Corley, Neil C.
TITLE OF INVENTION: DISEASE RELATED NUCLEOTIDE KINASES
NUMBER OF SEQUENCES: 12
CORRESPONDENCE ADDRESS:
ADDRESSEE: Incyte Pharmaceuticals, Inc.

STREET: 3174 Porter Drive
CITY: Palo Alto
STATE: CA
COUNTRY: USA
ZIP: 94304
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: FastSeq for Windows Version 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/879,561
FILING DATE: Herewith
CLASSIFICATION: 424
PRIOR APPLICATION DATA:
APPLICATION NUMBER:
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Billings, Lucy J.
REGISTRATION NUMBER: 36,749
REFERENCE/DOCKET NUMBER: PF-0325 US
TELECOMMUNICATION INFORMATION:
TELEPHONE: 415-855-0555
TELEFAX: 415-845-4166
TELEX:
INFORMATION FOR SEQ ID NO: 7:
SEQUENCE CHARACTERISTICS:
LENGTH: 260 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
IMMEDIATE SOURCE:
LIBRARY: GenBank
CLONE: 1480198
US-08-879-561-7

Query Match
Best Local Similarity 100.0%; Pred. No. 27; Length 260;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 245 QRAREEEE 252
Db 188 QRAREEEE 195

RESULT 11
US-08-979-608A-41
Sequence 41, Application US/08979608A
Patent No. 6355451
GENERAL INFORMATION:
APPLICANT: Lees, Ann M.
Lees, Robert S.
Law, Simon W.
Arjona, Anibal A.
TITLE OF INVENTION: NOVEL LOW DENSITY LIPOPROTEIN
BINDING PROTEINS AND THEIR USES IN DIAGNOSING AND
TREATING ATHEROSCLEROSIS
NUMBER OF SEQUENCES: 42
CORRESPONDENCE ADDRESS:
ADDRESSEE: Fish & Richardson P.C.
STREET: 225 Franklin Street
CITY: Boston
STATE: MA
COUNTRY: USA
ZIP: 02110-2804
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: FastSeq for Windows Version 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/979,608A
FILING DATE: 26-No. 6355451-1997

PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 60/048,547
FILING DATE: 03-JUN-1997
APPLICATION NUMBER: US 60/031,930
FILING DATE: 27-NOV-1996
ATTORNEY/AGENT INFORMATION:
NAME: Myers, Louis
REGISTRATION NUMBER: 35,965
REFERENCE/DOCKET NUMBER: 10797-002001 (formerly 3983/59818)
TELECOMMUNICATION INFORMATION:
TELEPHONE: 617/542-5070
TELEFAX: 617/542-8906
INFORMATION FOR SEQ ID NO: 41:
SEQUENCE CHARACTERISTICS:
LENGTH: 7 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
SEQUENCE DESCRIPTION: SEQ ID NO: 41:
US-08-979-608A-41

Query Match 1.3%; Score 7; DB 4; Length 7;
Best Local Similarity 100.0%; Pred. No. 3e+05; Indels 0; Gaps 0;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 53 RDVSEEL 59
Db 1 RDVSEEL 7

RESULT 12
US-09-517-849-41
Sequence 41, Application US/09517849
Patent No. 6605588
GENERAL INFORMATION:
APPLICANT: Lees, Ann M.
Lees, Robert S.
Law, Simon W.
Arjona, Anibal A.
TITLE OF INVENTION: NOVEL LOW DENSITY LIPOPROTEIN
BINDING PROTEINS AND THEIR USES IN DIAGNOSING AND
TREATING ATHEROSCLEROSIS
NUMBER OF SEQUENCES: 42
CORRESPONDENCE ADDRESS:
ADDRESSEE: Fish & Richardson P.C.
STREET: 225 Franklin Street
CITY: Boston
STATE: MA
COUNTRY: USA
ZIP: 02110-2804
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: FastSEQ for Windows Version 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/517,849
FILING DATE: 02-Mar-2000
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/979,608
FILING DATE: 26-NOV-1997
ATTORNEY/AGENT INFORMATION:
NAME: Myers, Louis
REGISTRATION NUMBER: 35,965
REFERENCE/DOCKET NUMBER: 10797-003001
TELECOMMUNICATION INFORMATION:
TELEPHONE: 617/542-5070
TELEFAX: 617/542-8906
INFORMATION FOR SEQ ID NO: 41:
SEQUENCE CHARACTERISTICS:
LENGTH: 7 amino acids
TYPE: amino acid
TOPOLOGY: linear

MOLECULE TYPE: protein
SEQUENCE DESCRIPTION: SEQ ID NO: 41:
US-09-517-849-41

Query Match 1.3%; Score 7; DB 4; Length 7;
Best Local Similarity 100.0%; Pred. No. 3e+05; Indels 0; Gaps 0;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 53 RDVSEEL 59
Db 1 RDVSEEL 7

RESULT 13
US-09-616-289-41
Sequence 41, Application US/09616289
Patent No. 6632923
GENERAL INFORMATION:
APPLICANT: Lees, Ann M.
Lees, Robert S.
Law, Simon W.
Arjona, Anibal A.
TITLE OF INVENTION: NOVEL LOW DENSITY LIPOPROTEIN BINDING
PROTEINS AND THEIR USE IN DIAGNOSING AND TREATING
ATHEROSCLEROSIS
FILE REFERENCE: 10797-004001
CURRENT APPLICATION NUMBER: US/09/616,289
CURRENT FILING DATE: 2000-07-14
PRIOR APPLICATION NUMBER: US 09/517,849
PRIOR FILING DATE: 2000-03-02
PRIOR APPLICATION NUMBER: US 08/979,608
PRIOR FILING DATE: 1997-11-26
PRIOR APPLICATION NUMBER: US 60/031,930
PRIOR FILING DATE: 1996-11-27
PRIOR APPLICATION NUMBER: US 60/048,547
PRIOR FILING DATE: 1997-06-03
NUMBER OF SEQ ID NOS: 53
SOFTWARE: FastSEQ for Windows Version 4.0
SEQ ID NO 41
LENGTH: 7
TYPE: PRT
ORGANISM: Homo sapiens
US-09-616-289-41

Query Match 1.3%; Score 7; DB 4; Length 7;
Best Local Similarity 100.0%; Pred. No. 3e+05; Indels 0; Gaps 0;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 53 RDVSEEL 59
Db 1 RDVSEEL 7

RESULT 14
US-09-146-950-25
Sequence 25, Application US/09146950A
Patent No. 6287808
GENERAL INFORMATION:
APPLICANT: Busfield, Samantha J.
TITLE OF INVENTION: NOVEL MOLECULES OF THE HERPESVIRUS-ENTRY-MEDIATOR-RELATED
FAMILY AND USES THEREOF
FILE REFERENCE: 09404/057001
CURRENT APPLICATION NUMBER: US/09/146,950A
CURRENT FILING DATE: 1998-09-03
NUMBER OF SEQ ID NOS: 25
SOFTWARE: FastSEQ for Windows Version 3.0
SEQ ID NO 25
LENGTH: 77
TYPE: PRT
ORGANISM: Homo sapiens
US-09-146-950-25

Query Match 1.3%; Score 7; DB 3; Length 77;

Best Local Similarity 100.0%; Pred. No. 82;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 145 EKKKAG 151
|||||
Db 71 EKKKAG 77

RESULT 15
US-09-732-210-434
; Sequence 434, Application US/09732210
; Patent No. 6573361
; GENERAL INFORMATION:
; APPLICANT: Bunkers, Greg J.
; APPLICANT: Lians, Jihong
; APPLICANT: Mittanck, Cindy A.
; APPLICANT: Seale, Jeffrey W.
; APPLICANT: Wu, Yomie S.
; TITLE OF INVENTION: Anti-fungal Proteins and Methods for Their Use
; FILE REFERENCE: 38-21(15036)B
; CURRENT APPLICATION NUMBER: US/09/732,210
; CURRENT FILING DATE: 2000-12-07
; PRIOR APPLICATION NUMBER: US 60/169,513
; PRIOR FILING DATE: 1999-12-07
; PRIOR APPLICATION NUMBER: US 60/169,340
; PRIOR FILING DATE: 1999-12-07
; NUMBER OF SEQ ID NOS: 1753
; SEQ ID NO 434
; LENGTH: 105
; TYPE: PRT
; ORGANISM: Odontella sinensis
US-09-732-210-434

Query Match 1.3%; Score 7; DB 4; Length 105;
Best Local Similarity 100.0%; Pred. No. 1.1e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 152 LGKEITL 158
|||||
Db 30 LGKEITL 36

RESULT 16
US-09-489-039A-10493
; Sequence 10493, Application US/09489039A
; Patent No. 6610836
; GENERAL INFORMATION:
; APPLICANT: Gary Breton et. al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO KLEBSIELLA
; TITLE OF INVENTION: PNEUMONIAE FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 2709.2004001
; CURRENT APPLICATION NUMBER: US/09/489,039A
; CURRENT FILING DATE: 2000-01-27
; PRIOR APPLICATION NUMBER: US 60/117,747
; PRIOR FILING DATE: 1999-01-29
; NUMBER OF SEQ ID NOS: 14342
; SEQ ID NO 10493
; LENGTH: 124
; TYPE: PRT
; ORGANISM: Klebsiella pneumoniae
US-09-489-039A-10493

Query Match 1.3%; Score 7; DB 4; Length 124;
Best Local Similarity 100.0%; Pred. No. 1.2e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 346 LLKEAVE 352
|||||
Db 111 LLKEAVE 117

RESULT 17
US-08-849-303-19

; Sequence 19, Application US/08849303
; Patent No. 6680424
; GENERAL INFORMATION:
; APPLICANT: Ackinson, Howard J.
; APPLICANT: McPherson, Michael J.
; APPLICANT: Urwin, Peter E.
; TITLE OF INVENTION: MODIFIED PROTEINASE INHIBITORS
; NUMBER OF SEQUENCES: 79
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Klauber & Jackson
; STREET: 411 Hackensack Avenue, 4th Floor
; CITY: Hackensack
; STATE: New Jersey
; COUNTRY: USA
; ZIP: 07601
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/849,303
; FILING DATE: 21-MAY-1997
; CLASSIFICATION: 514
; ATTORNEY/AGENT INFORMATION:
; NAME: Jackson Esq., David A.
; REGISTRATION NUMBER: 26,742
; REFERENCE/DOCKET NUMBER: 1321-1-003
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 201-487-5800
; TELEFAX: 201-343-1684
; TELEX: 133521
; INFORMATION FOR SEQ ID NO: 19:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 127 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; HYPOTHETICAL: NO
US-08-849-303-19

Query Match 1.3%; Score 7; DB 4; Length 127;
Best Local Similarity 100.0%; Pred. No. 1.2e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 241 BEGVORA 247
|||||
Db 27 BEGVORA 33

RESULT 18
US-09-489-039A-9418
; Sequence 9418, Application US/09489039A
; Patent No. 6610836
; GENERAL INFORMATION:
; APPLICANT: Gary Breton et. al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO KLEBSIELLA
; TITLE OF INVENTION: PNEUMONIAE FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 2709.2004001
; CURRENT APPLICATION NUMBER: US/09/489,039A
; CURRENT FILING DATE: 2000-01-27
; PRIOR APPLICATION NUMBER: US 60/117,747
; PRIOR FILING DATE: 1999-01-29
; NUMBER OF SEQ ID NOS: 14342
; SEQ ID NO 9418
; LENGTH: 142
; TYPE: PRT
; ORGANISM: Klebsiella pneumoniae
US-09-489-039A-9418

Query Match 1.3%; Score 7; DB 4; Length 142;
Best Local Similarity 100.0%; Pred. No. 1.4e+02;

Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 157 TLLMQTL 163
Db 60 TLLMQTL 66

RESULT 19
US-09-146-950-20
; Sequence 20, Application US/09146950A
; Patent No. 6287808
; GENERAL INFORMATION:
; APPLICANT: Busfield, Samantha J.
; TITLE OF INVENTION: NOVEL MOLECULES OF THE HERPESVIRUS-ENTRY-MEDIATOR-RELATED
; FILE REFERENCE: 09404/057001
; CURRENT APPLICATION NUMBER: US/09/146,950A
; CURRENT FILING DATE: 1998-09-03
; NUMBER OF SEQ ID NOS: 25
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 20
; LENGTH: 159
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-146-950-20

Query Match 1.3%; Score 7; DB 3; Length 159;
Best Local Similarity 100.0%; Pred. No. 1.5e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 145 EKKKAGK 151
Db 153 EKKKAGK 159

RESULT 20
US-09-079-030-117
; Sequence 117, Application US/09079030
; Patent No. 6635623
; GENERAL INFORMATION:
; APPLICANT: Guevera, Jr., Juan G.
; APPLICANT: Hoogveen, Ron C.
; APPLICANT: Moore, Paul J.
; TITLE OF INVENTION: LIPOPROTEINS AS NUCLEIC ACID DELIVERY
; TITLE OF INVENTION: VECTORS FOR TRANSFECTION OF EUKARYOTIC CELLS
; NUMBER OF SEQUENCES: 229
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Arnold, White & Durkee
; STREET: P.O. Box 4433
; CITY: Houston
; STATE: Texas
; COUNTRY: USA
; ZIP: 77210
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/079,030
; FILING DATE: Concurrently Herewith
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: McMillian, Nabeela R.
; REGISTRATION NUMBER: P-43,363
; REFERENCE/DOCKET NUMBER: ARAG:003
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 512/418-3000
; TELEFAX: 512/474-7577
; INFORMATION FOR SEQ ID NO: 117:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 172 amino acids
; TYPE: amino acid

STRANDEDNESS:
TOPOLOGY: linear
US-09-079-030-117

Query Match 1.3%; Score 7; DB 4; Length 172;
Best Local Similarity 100.0%; Pred. No. 1.6e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 281 KLRQKNM 287
Db 100 KLRQKNM 106

RESULT 21
US-09-328-352-7281
; Sequence 7281, Application US/09328352
; Patent No. 6562958
; GENERAL INFORMATION:
; APPLICANT: Gary L. Breton et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO ACINETOBACTER
; TITLE OF INVENTION: BAUMANNII FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: GTC99-03PA
; CURRENT APPLICATION NUMBER: US/09/328,352
; CURRENT FILING DATE: 1999-06-04
; NUMBER OF SEQ ID NOS: 8252
; SEQ ID NO 7281
; LENGTH: 187
; TYPE: PRT
; ORGANISM: Acinetobacter baumannii
US-09-328-352-7281

Query Match 1.3%; Score 7; DB 4; Length 187;
Best Local Similarity 100.0%; Pred. No. 1.7e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 439 KELEGLQ 445
Db 83 KELEGLQ 89

RESULT 22
US-09-146-950-18
; Sequence 18, Application US/09146950A
; Patent No. 6287808
; GENERAL INFORMATION:
; APPLICANT: Busfield, Samantha J.
; TITLE OF INVENTION: NOVEL MOLECULES OF THE HERPESVIRUS-ENTRY-MEDIATOR-RELATED
; TITLE OF INVENTION: PROTEIN FAMILY AND USES THEREOF
; FILE REFERENCE: 09404/057001
; CURRENT APPLICATION NUMBER: US/09/146,950A
; CURRENT FILING DATE: 1998-09-03
; NUMBER OF SEQ ID NOS: 25
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 18
; LENGTH: 197
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-146-950-18

Query Match 1.3%; Score 7; DB 3; Length 197;
Best Local Similarity 100.0%; Pred. No. 1.8e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 145 EKKKAGK 151
Db 191 EKKKAGK 197

RESULT 23
US-09-134-001C-5175
; Sequence 5175, Application US/09134001C
; Patent No. 6380370
; GENERAL INFORMATION:

APPLICANT: Lynn Doucette-Stamm et al
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS
FILE REFERENCE: GTC-007
CURRENT APPLICATION NUMBER: US/09/134,001C
PRIOR FILING DATE: 1998-08-13
PRIOR APPLICATION NUMBER: US 60/064,964
PRIOR FILING DATE: 1997-11-08
PRIOR APPLICATION NUMBER: US 60/055,779
PRIOR FILING DATE: 1997-08-14
NUMBER OF SEQ ID NOS: 5674
SEQ ID NO 5175
LENGTH: 211
TYPE: PRT
ORGANISM: Staphylococcus epidermidis
US-09-134-001C-5175

Query Match 1.3%; Score 7; DB 4; Length 211;
Best Local Similarity 100.0%; Pred. No. 1.9e+02; Indels 0; Gaps 0;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 404 KKKKLE 410
Db 19 KKKKLE 25

RESULT 24
US-08-879-561-1
Sequence 1, Application US/08879561
Patent No. 5817482
GENERAL INFORMATION:
APPLICANT: Bandman, Olga
APPLICANT: Hillman, Jennifer L.
APPLICANT: Hawkins, Phillip R.
APPLICANT: Guegler, Karl J.
APPLICANT: Corley, Neil C.
TITLE OF INVENTION: DISEASE RELATED NUCLEOTIDE KINASES
NUMBER OF SEQUENCES: 12
CORRESPONDENCE ADDRESS:
ADDRESSEE: Incyte Pharmaceuticals, Inc.
STREET: 3174 Porter Drive
CITY: Palo Alto
STATE: CA
COUNTRY: USA
ZIP: 94304
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: FASTSEQ for Windows Version 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/879,561
FILING DATE: Herewith
CLASSIFICATION: 424
PRIOR APPLICATION DATA:
APPLICATION NUMBER:
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Billings, Lucy J.
REGISTRATION NUMBER: 36,749
REFERENCE/DOCKET NUMBER: PF-0325 US
TELEPHONE: 415-855-0555
TELEFAX: 415-845-4166
TELEX:
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 261 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
IMMEDIATE SOURCE:
LIBRARY: FIBROT01

APPLICANT: GARY BRETON et al
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO KLEBSIELLA
FILE REFERENCE: 2709.20040001
CURRENT APPLICATION NUMBER: US/09/489,039A
PRIOR FILING DATE: 2000-01-27
PRIOR APPLICATION NUMBER: US 60/117,747
PRIOR FILING DATE: 1999-01-29
NUMBER OF SEQ ID NOS: 14342
SEQ ID NO 13457
LENGTH: 267
TYPE: PRT
ORGANISM: Klebsiella pneumoniae
US-09-489-039A-13457

Query Match 1.3%; Score 7; DB 4; Length 267;
Best Local Similarity 100.0%; Pred. No. 2.3e+02; Indels 0; Gaps 0;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 493 GAQAPSS 499
Db 113 GAQAPSS 119

RESULT 26
US-08-287-959-3
Sequence 3, Application US/08287959
Patent No. 5639651
GENERAL INFORMATION:
APPLICANT: Weisbach, Lawrence
APPLICANT: Bernards, Andre
APPLICANT: Settleman, Jeffrey
TITLE OF INVENTION: GAP-RELATED GENE
NUMBER OF SEQUENCES: 26
CORRESPONDENCE ADDRESS:
ADDRESSEE: Fish & Richardson
STREET: 225 Franklin Street
CITY: Boston
STATE: MA
COUNTRY: U.S.A.
ZIP: 02110
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/287,959
FILING DATE: August 9, 1994
CLASSIFICATION: 514
ATTORNEY/AGENT INFORMATION:
NAME: Clark, Paul C.
REGISTRATION NUMBER: 30,162
REFERENCE/DOCKET NUMBER: 00786/181001
TELECOMMUNICATION INFORMATION:
TELEPHONE: (617) 542-5070

TELEFAX: (617) 542-8906
TELEX: 200154
INFORMATION FOR SEQ ID NO: 3:
SEQUENCE CHARACTERISTICS:
LENGTH: 274 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: peptide
US-08-287-959-3

Query Match 1.3%; Score 7; DB 1; Length 274;
Best Local Similarity 100.0%; Pred. No. 2.3e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 471 DLSAGGQ 477
Db 213 DLSAGGQ 219

RESULT 27
US-09-857-498A-7
Sequence 7, Application US/09857498A
Patent No. 6582947
GENERAL INFORMATION:
APPLICANT: Karlsson, Anna
APPLICANT: Johansson, Magnus
TITLE OF INVENTION: New medical use
FILE REFERENCE: multibstrate deoxyguanosine kinase
CURRENT APPLICATION NUMBER: US/09/857,498A
CURRENT FILING DATE: 2002-01-17
PRIOR APPLICATION NUMBER: SE 9804298-9
PRIOR FILING DATE: 1998-12-11
NUMBER OF SEQ ID NOS: 8
SOFTWARE: PatentIn Ver. 2.1
SEQ ID NO 7
LENGTH: 277
TYPE: PRT
ORGANISM: Homo sapiens
US-09-857-498A-7

Query Match 1.3%; Score 7; DB 4; Length 277;
Best Local Similarity 100.0%; Pred. No. 2.3e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 245 QRAREEE 251
Db 205 QRAREEE 211

RESULT 28
US-09-134-000C-3980
Sequence 3980, Application US/09134000C
Patent No. 6617156
GENERAL INFORMATION:
APPLICANT: Lynn Doucette-Stamm et al
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO
FILE REFERENCE: 032796-032
CURRENT APPLICATION NUMBER: US/09/134,000C
CURRENT FILING DATE: 1998-08-13
PRIOR APPLICATION NUMBER: US 60/055,778
PRIOR FILING DATE: 1997-08-15
NUMBER OF SEQ ID NOS: 6812
SOFTWARE: PatentIn version 3.1
SEQ ID NO 3980
LENGTH: 281
TYPE: PRT
ORGANISM: Enterococcus faecalis
US-09-134-000C-3980

Query Match 1.3%; Score 7; DB 4; Length 281;
Best Local Similarity 100.0%; Pred. No. 2.4e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 224 KLESCLR 230
Db 50 KLESCLR 56

RESULT 29
US-09-489-039A-8327
Sequence 8327, Application US/09489039A
Patent No. 6610836
GENERAL INFORMATION:
APPLICANT: Gary Breton et. al
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO KLEBSIELLA
FILE REFERENCE: PNEUMONIAE FOR DIAGNOSTICS AND THERAPEUTICS
CURRENT APPLICATION NUMBER: US/09/489,039A
CURRENT FILING DATE: 2000-01-27
PRIOR APPLICATION NUMBER: US 60/117,747
PRIOR FILING DATE: 1999-01-29
NUMBER OF SEQ ID NOS: 14342
SEQ ID NO 8327
LENGTH: 294
TYPE: PRT
ORGANISM: Klebsiella pneumoniae
US-09-489-039A-8327

Query Match 1.3%; Score 7; DB 4; Length 294;
Best Local Similarity 100.0%; Pred. No. 2.5e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 182 ELLEHR 188
Db 19 ELLEHR 25

RESULT 30
US-09-328-352-4508
Sequence 4508, Application US/09328352
Patent No. 6562958
GENERAL INFORMATION:
APPLICANT: Gary L. Breton et al.
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO ACINETOBACTER
FILE REFERENCE: BAUMANNII FOR DIAGNOSTICS AND THERAPEUTICS
CURRENT APPLICATION NUMBER: US/09/328,352
CURRENT FILING DATE: 1999-06-04
NUMBER OF SEQ ID NOS: 8252
SEQ ID NO 4508
LENGTH: 302
TYPE: PRT
ORGANISM: Acinetobacter baumannii
US-09-328-352-4508

Query Match 1.3%; Score 7; DB 4; Length 302;
Best Local Similarity 100.0%; Pred. No. 2.5e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 345 FILKEAV 351
Db 102 FILKEAV 108

RESULT 31
US-09-252-991A-28792
Sequence 28792, Application US/09252991A
Patent No. 6551795
GENERAL INFORMATION:
APPLICANT: Marc J. Rubenfield et al.
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
FILE REFERENCE: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
CURRENT APPLICATION NUMBER: US/09/252,991A
CURRENT FILING DATE: 1999-02-18

; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 28792
; LENGTH: 311
; TYPE: PRT
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-28792

Query Match 1.3%; Score 7; DB 4; Length 311;
Best Local Similarity 100.0%; Pred. No. 2.6e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 131 QSDVEGD 137
Db 148 QSDVEGD 154

RESULT 32
US-08-285-440-4
; Sequence 4, Application US/08285440
; Patent No. 5532337
; GENERAL INFORMATION:
; APPLICANT: Ken'ichiro HAYASHI et al.
; TITLE OF INVENTION: POLYPEPTIDE
; NUMBER OF SEQUENCES: 32
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Wenderoth, Lind & Ponack
; STREET: 805 Fifteenth Street, N.W., #700
; CITY: Washington
; STATE: D.C.
; COUNTRY: U.S.A.
; ZIP: 20005
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 5.25 inch, 500 kb
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: MS-DOS
; SOFTWARE: DisplayWrite
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/285,440
; FILING DATE:
; CLASSIFICATION: 536
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 07/858,947
; FILING DATE: March 27, 1992
; ATTORNEY/AGENT INFORMATION:
; NAME: Warren M. Cheek, Jr.
; REGISTRATION NUMBER: 33,367
; REFERENCE/DOCKET NUMBER:
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 202-371-8850
; TELEFAX:
; TELEX:
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 312 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; HYPOTHEetical:
; ANTI-SENSE:
; FRAGMENT TYPE:
; ORIGINAL SOURCE:
; ORGANISM:
; STRAIN:
; INDIVIDUAL ISOLATE:
; DEVELOPMENTAL STAGE:
; HAPLOTYPE:
; TISSUE TYPE:
; CELL TYPE:

; CELL LINE:
; ORGANELL:
; IMMEDIATE SOURCE:
; LIBRARY:
; CLONE:
; POSITION IN GENOME:
; CHROMOSOME/SEGMENT:
; MAP POSITION:
; UNITS:
; FEATURE:
; NAME/KEY:
; LOCATION:
; IDENTIFICATION METHOD:
; OTHER INFORMATION:
; PUBLICATION INFORMATION:
; AUTHORS:
; TITLE:
; JOURNAL:
; VOLUME:
; ISSUE:
; PAGES:
; DATE:
; DOCUMENT NUMBER:
; FILING DATE:
; PUBLICATION DATE:
; RELEVANT RESIDUES IN SEQ ID NO:
US-08-285-440-4

Query Match 1.3%; Score 7; DB 1; Length 312;
Best Local Similarity 100.0%; Pred. No. 2.6e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 248 REEEKR 254
Db 115 REEEKR 121

RESULT 33
US-08-630-349-4
; Sequence 4, Application US/08630349
; Patent No. 5739008
; GENERAL INFORMATION:
; APPLICANT: Ken'ichiro HAYASHI et al.
; TITLE OF INVENTION: POLYPEPTIDE
; NUMBER OF SEQUENCES: 32
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Wenderoth, Lind & Ponack
; STREET: 805 Fifteenth Street, N.W., #700
; CITY: Washington
; STATE: D.C.
; COUNTRY: U.S.A.
; ZIP: 20005
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.5 inch, 144 mb
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: MS-DOS
; SOFTWARE: Wordperfect 5.1
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/630,349
; FILING DATE: April 10, 1996
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/285,440
; FILING DATE: August 4, 1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 07/858,947
; FILING DATE: March 27, 1992
; ATTORNEY/AGENT INFORMATION:
; NAME: Warren M. Cheek, Jr.
; REGISTRATION NUMBER: 33,367
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 312 amino acids

; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
US-08-630-349-4

Query Match 1.3%; Score 7; DB 1; Length 312;
Best Local Similarity 100.0%; Pred. No. 2.6e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 248 REEEKR 254
Db 115 REEEKR 121

RESULT 34

US-09-252-991A-18599
; Sequence 18599, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 18599
; LENGTH: 314
; TYPE: PRT
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-18599

Query Match 1.3%; Score 7; DB 4; Length 314;
Best Local Similarity 100.0%; Pred. No. 2.6e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 131 QSDVEGD 137
Db 147 QSDVEGD 153

RESULT 35

US-09-252-991A-28531
; Sequence 28531, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 28531
; LENGTH: 330
; TYPE: PRT
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-28531

Query Match 1.3%; Score 7; DB 4; Length 330;
Best Local Similarity 100.0%; Pred. No. 2.7e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 138 RDHRRPQ 144

Db 259 RDHRRPQ 265

RESULT 36

US-09-008-465-1
; Sequence 1, Application US/09008465
; Patent No. 6174702
; GENERAL INFORMATION:
; APPLICANT: Lal, Preeti
; APPLICANT: Guegler, Karl J.
; APPLICANT: Corley, Neil C.
; TITLE OF INVENTION: HUMAN PINCH PROTEIN HOMOLOG
; NUMBER OF SEQUENCES: 3
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Incyte Pharmaceuticals, Inc.
; STREET: 3174 Porter Dr.
; CITY: Palo Alto
; STATE: CA
; COUNTRY: USA
; ZIP: 94304
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/008,465
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Billings, Lucy J
; REGISTRATION NUMBER: 36,749
; REFERENCE/DOCKET NUMBER: PF-0460 US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 650-855-0555
; TELEFAX: 650-845-4166
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 341 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; IMMEDIATE SOURCE:
; LIBRARY: SEMVNOT04
; CLONE: 3540806
US-09-008-465-1

Query Match 1.3%; Score 7; DB 3; Length 341;
Best Local Similarity 100.0%; Pred. No. 2.8e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 148 KAKGLGK 154
Db 131 KAKGLGK 137

RESULT 37

US-09-528-959-1
; Sequence 1, Application US/09528959
; Patent No. 6379904
; GENERAL INFORMATION:
; APPLICANT: Lal, Preeti
; APPLICANT: Guegler, Karl J.
; APPLICANT: Corley, Neil C.
; TITLE OF INVENTION: HUMAN PINCH PROTEIN HOMOLOG
; NUMBER OF SEQUENCES: 3
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Incyte Pharmaceuticals, Inc.
; STREET: 3174 Porter Dr.

CITY: Palo Alto
STATE: CA
COUNTRY: USA
ZIP: 94304
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: FASTSEQ for Windows Version 2.0
CURRENT APPLICATION DATA: US/09/528,959
FILING DATE: 20-Mar-2000
CLASSIFICATION: <Unknown>
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 09/008,465
FILING DATE: <Unknown>
ATTORNEY/AGENT INFORMATION:
NAME: Billings, Lucy J
REGISTRATION NUMBER: 36,749
REFERENCE/DOCKET NUMBER: PF-0460 US
TELECOMMUNICATION INFORMATION:
TELEPHONE: 650-855-0555
TELEFAX: 650-845-4166
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 341 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
IMMEDIATE SOURCE:
LIBRARY: SEMWNOT04
CLONE: 3540806
SEQUENCE DESCRIPTION: SEQ ID NO: 1:
US-09-528-959-1

Query Match 1.3%; Score 7; DB 4; Length 341;
Best Local Similarity 100.0%; Pred. No. 2.8e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 148 KAKGLGK 154
Db 131 KAKGLGK 137
RESULT 38
US-09-252-991A-27167
; Sequence 27167, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; PRIOR FILING DATE: 1999-02-18
; PRIOR FILING DATE: 1998-02-18
; PRIOR FILING DATE: 1998-02-18
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 27167
; LENGTH: 346
; TYPE: PRT
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-27167

Query Match 1.3%; Score 7; DB 4; Length 346;
Best Local Similarity 100.0%; Pred. No. 2.8e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 21 AAPAVEA 27
Db 62 AAPAVEA 69
RESULT 39
US-09-252-991A-27209
; Sequence 27209, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; PRIOR FILING DATE: 1999-02-18
; PRIOR FILING DATE: 1998-02-18
; PRIOR FILING DATE: 1998-02-18
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 27209
; LENGTH: 364
; TYPE: PRT
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-27209

Query Match 1.3%; Score 7; DB 4; Length 388;
Best Local Similarity 100.0%; Pred. No. 3.1e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 487 RRPECPG 493
Db 76 RRPECPG 82
RESULT 41
US-09-543-681A-7318
; Sequence 7318, Application US/09543681A
; Patent No. 6605709
; GENERAL INFORMATION:
; APPLICANT: GARY BRITON
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PROTEUS MIRABIL
; FILE REFERENCE: 2709.1002-001

RESULT 39
US-09-252-991A-27209
; Sequence 27209, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; PRIOR FILING DATE: 1999-02-18
; PRIOR FILING DATE: 1998-02-18
; PRIOR FILING DATE: 1998-02-18
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 27209
; LENGTH: 364
; TYPE: PRT
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-27209
Query Match 1.3%; Score 7; DB 4; Length 364;
Best Local Similarity 100.0%; Pred. No. 2.9e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 136 GDRDHR 142
Db 332 GDRDHR 338

RESULT 40
US-09-252-991A-21572
; Sequence 21572, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; PRIOR FILING DATE: 1999-02-18
; PRIOR FILING DATE: 1998-02-18
; PRIOR FILING DATE: 1998-02-18
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 21572
; LENGTH: 388
; TYPE: PRT
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-21572

Query Match 1.3%; Score 7; DB 4; Length 388;
Best Local Similarity 100.0%; Pred. No. 3.1e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 487 RRPECPG 493
Db 76 RRPECPG 82
RESULT 41
US-09-543-681A-7318
; Sequence 7318, Application US/09543681A
; Patent No. 6605709
; GENERAL INFORMATION:
; APPLICANT: GARY BRITON
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PROTEUS MIRABIL
; FILE REFERENCE: 2709.1002-001

; CURRENT APPLICATION NUMBER: US/09/543,681A
; CURRENT FILING DATE: 2000-04-05
; PRIOR APPLICATION NUMBER: US 60/128,706
; PRIOR FILING DATE: 1999-04-09
; NUMBER OF SEQ ID NOS: 8344
; SEQ ID NO 7318
; LENGTH: 389
; TYPE: PRT
; ORGANISM: Proteus mirabilis
US-09-543-681A-7318

Query Match 1.3%; Score 7; DB 4; Length 389;
Best Local Similarity 100.0%; Pred. No. 3.1e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 109 PTPVVG 115
Db 329 PTPVVG 335
|||||

RESULT 42
US-09-252-991A-24632
; Sequence 24632, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 24632
; LENGTH: 411
; TYPE: PRT
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-24632

; CURRENT APPLICATION NUMBER: US/09/543,681A
; CURRENT FILING DATE: 2000-04-05
; PRIOR APPLICATION NUMBER: US 60/128,706
; PRIOR FILING DATE: 1999-04-09
; NUMBER OF SEQ ID NOS: 8344
; SEQ ID NO 7318
; LENGTH: 389
; TYPE: PRT
; ORGANISM: Proteus mirabilis
US-09-543-681A-7318

Query Match 1.3%; Score 7; DB 4; Length 411;
Best Local Similarity 100.0%; Pred. No. 3.2e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 4 PQQPEAG 10
Db 149 PQQPEAG 155
|||||

RESULT 43
US-09-252-991A-21571
; Sequence 21571, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 21571
; LENGTH: 422
; TYPE: PRT
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-21571

; CURRENT APPLICATION NUMBER: US/09/543,681A
; CURRENT FILING DATE: 2000-04-05
; PRIOR APPLICATION NUMBER: US 60/128,706
; PRIOR FILING DATE: 1999-04-09
; NUMBER OF SEQ ID NOS: 8344
; SEQ ID NO 7318
; LENGTH: 389
; TYPE: PRT
; ORGANISM: Proteus mirabilis
US-09-543-681A-7318

Query Match 1.3%; Score 7; DB 4; Length 411;
Best Local Similarity 100.0%; Pred. No. 3.2e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 4 PQQPEAG 10
Db 149 PQQPEAG 155
|||||

RESULT 43
US-09-252-991A-21571
; Sequence 21571, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 21571
; LENGTH: 422
; TYPE: PRT
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-21571

Query Match 1.3%; Score 7; DB 4; Length 422;
Best Local Similarity 100.0%; Pred. No. 3.3e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 487 RRPEGPG 493
Db 16 RRPEGPG 22
|||||

RESULT 44
US-09-252-991A-25318
; Sequence 25318, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 25318
; LENGTH: 446
; TYPE: PRT
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-25318

; CURRENT APPLICATION NUMBER: US/09/543,681A
; CURRENT FILING DATE: 2000-04-05
; PRIOR APPLICATION NUMBER: US 60/128,706
; PRIOR FILING DATE: 1999-04-09
; NUMBER OF SEQ ID NOS: 8344
; SEQ ID NO 7318
; LENGTH: 389
; TYPE: PRT
; ORGANISM: Proteus mirabilis
US-09-543-681A-7318

Query Match 1.3%; Score 7; DB 4; Length 446;
Best Local Similarity 100.0%; Pred. No. 3.5e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 287 MELAEEL 293
Db 415 MELAEEL 421
|||||

RESULT 45
US-09-252-991A-16904
; Sequence 16904, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 16904
; LENGTH: 473
; TYPE: PRT
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-16904

; CURRENT APPLICATION NUMBER: US/09/543,681A
; CURRENT FILING DATE: 2000-04-05
; PRIOR APPLICATION NUMBER: US 60/128,706
; PRIOR FILING DATE: 1999-04-09
; NUMBER OF SEQ ID NOS: 8344
; SEQ ID NO 7318
; LENGTH: 389
; TYPE: PRT
; ORGANISM: Proteus mirabilis
US-09-543-681A-7318

Query Match 1.3%; Score 7; DB 4; Length 473;
Best Local Similarity 100.0%; Pred. No. 3.6e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 41 GAQARTA 47
Db 418 GAQARTA 424
|||||

RESULT 46
US-09-543-681A-6752


```

RA Chikuni K., Tanabe R., Muroya S., Nakajima I.;
RT "Differences in molecular structure among the porcine myosin heavy
RL chain-2a, -2x, and -2b isoforms.";
DR Meat Sci. 57:311-317(2001).
DR EMBL; AB025261; BAA82145.1; -.
DR HSSP; P13538; 2MYS.
DR GO; GO:0016459; C:myosin; IEA.
DR GO; GO:0005524; F:ATP binding; IEA.
DR GO; GO:0003774; F:motor activity; IEA.
DR InterPro; IPR000048; IQ_region.
DR InterPro; IPR001609; myosin_head.
DR InterPro; IPR004009; Myosin_N.
DR InterPro; IPR002928; Myosin_tail.
DR Pfam; PF00612; IQ; 2.
DR Pfam; PF00663; myosin_head; 1.
DR Pfam; PF02736; Myosin_N; 1.
DR Pfam; PF01576; Myosin_tail; 1.
DR PRINTS; PR00193; MYOSINHEAVY.
DR ProDom; PD000355; myosin_head; 1.
DR SMART; SM00015; IQ; 1.
DR SMART; SM00242; MYSC; 1.
DR PROSITE; PS50096; IQ; 1.
SQ SEQUENCE 1937 AA; 223235 MW; BBC114C6824E0426 CRC64;

Query Match 9.7%; Score 262; DB 6; Length 1937;
Best Local Similarity 21.9%; Pred. No. 0.00015;
Matches 118; Conservative 78; Mismatches 172; Indels 170; Gaps 15;

Qy 25 VEAEGPSSAPKPEGAQRTAQSGALRDVSELSQLEDILSTYCVNNQGGPGEDGA 84
Db 1096 IEDQALAMQLKKIKELQART-----BELLEEIE----- 1125
Qy 85 QGPAEPEDAEKSRITYVARGPEPEPTPVVYGEKPSKGDPTNTEIROSQDVGDRHRRPQ 144
Db 1126 -AFASPAKAEKQSDLSRELEB-----ISERLEAGGATSAQI-----EMNKKREA 1171
Qy 145 EKKAKGLGHEITLLMOTLNTLSTPEKLAALCKY-----AELEHNSQKQKMLLOKK 200
Db 1172 EFGQREDELEATL-----CHEATAAALRKHADSVAEELGEQIDNLRVQKLEKE 1222
Qy 201 QSOLVQEKDHLRGHSAVLARSKLESICREL-----QRNRSIKKEGVQAR- 248
Db 1223 KSELKMEIDDLASNETVSKAGNLEKMCRTLEDQLESEVTKKEHQRLNELSAQKARL 1282
Qy 249 -----EEERKKEVTSH-FQVTLNDI 268
Db 1283 QTESGFSRQDKEALVLSQSRGKAFTQOIIEELKRLKEETKAKSALAHAVQSSRHDC 1342
Qy 269 QIQMEQHNER-----NSKLR-----QENMELAERLKCLIEQVELR 303
Db 1343 DLUREQYEESEQAELQRAMSKANSEVAQWRKYETDAIQTREELEBAKKLAQRLQDA 1402
Qy 304 EEHIDKVPKHKDQQQVDAKLAQAQAEMLKEAEERHQEKDFLLKEAVESQRMCELMKQ 363
Db 1403 EEHV-----EAVNAKAS-----LEKTKQLQNEVEDMLDVERSNAAACAALDKK 1447
Qy 364 ETHLKQOALYTKPEFEPQNTLSKSEVFTFPKQEMKMTKKIKLEKETMYRSWESS 423
Db 1448 QNPKDILAEKMKHYEETQ-----AELEASQKESRSLSSTELFKVKNAYEES 1493
Qy 424 NKALLEAAEEKTVRDKLEGLQVK-----IQRLKLCRALQTERNDLNKRVDLSA 474
Db 1494 LDQLETAKRENKNIQQHSIDLTQEIABGGKHIELEKVKQIQEKKSELQAALAEBA 1551

RESULT 30
Q9DGM4 PRELIMINARY; PRT; 1941 AA.
AC Q9DGM4;
DT 01-MAR-2001 (TrEMBLrel. 16, Created)
DT 01-MAR-2001 (TrEMBLrel. 16, last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, last annotation update)
DE Fast myosin heavy chain isoform 3.

```

```

OS Gallus gallus (Chicken).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;
OC Gallus.
OX NCBI_TaxID=9031;
RN [1]
RP SEQUENCE FROM N.A.
RA Zhang Q., Bandman E.;
RT "Seven skeletal myosin heavy chain genes (MyHC) are organized as a
RT multigene complex in the chicken genome.";
RL Submitted (MAY-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF272034; AAF99315.1; -.
DR HSSP; P13538; 2MYS.
DR GO; GO:0016459; C:myosin; IEA.
DR GO; GO:0005524; F:ATP binding; IEA.
DR GO; GO:0003774; F:motor activity; IEA.
DR InterPro; IPR000048; IQ_region.
DR InterPro; IPR001609; myosin_head.
DR InterPro; IPR004009; Myosin_N.
DR InterPro; IPR002928; Myosin_tail.
DR Pfam; PF00612; IQ; 2.
DR Pfam; PF00663; myosin_head; 1.
DR Pfam; PF02736; Myosin_N; 1.
DR Pfam; PF01576; Myosin_tail; 1.
DR PRINTS; PR00193; MYOSINHEAVY.
DR ProDom; PD000355; myosin_head; 1.
DR SMART; SM00015; IQ; 1.
DR SMART; SM00242; MYSC; 1.
DR PROSITE; PS50096; IQ; 1.
SQ SEQUENCE 1941 AA; 223162 MW; 9C8597C1CCTFDBEC CRC64;

Query Match 9.6%; Score 260; DB 13; Length 1941;
Best Local Similarity 22.6%; Pred. No. 0.00019;
Matches 118; Conservative 82; Mismatches 181; Indels 142; Gaps 16;

Qy 33 SQAPKPEGAQRTAQSGALRDVSELSQLEDILSTYCVNNQGGPGEDGAQGEPAEPE 92
Db 1093 SQIQSKIEDEALQWQ---LQKKIKELQARIEL-----EEIEAETSRA 1135
Qy 93 DAEKSRITYVARGPEPEPTPVVYGEKPSKGDPTNTEIROSQDVGDRHRRPQKKAAGL 152
Db 1136 KAEKHRADLSRELEB-----ISERLEAGGATAAQI-----DMNKKREAEPQWRD 1182
Qy 153 GHEITLLMOTLNTLSTPEKLAALCKY-----AELEHNSQKQKMLLOKQSLVQEK 208
Db 1183 LEETL-----CHEATAAALRKHADSTAEELGEQIDNLRVQKLEKSELSAMEI 1233
Qy 209 DHLRGHSAVLARSKLESICREL-----QRNRSIKKEGVQAR----- 248
Db 1234 DLLASNWSVSXKANKLEKMCRTLEDQLESEVTKKEHQRMINDLNTQARLQTEAGYS 1293
Qy 249 -----EEERKKEVTSH-FQVTLNDIQLQMEQHN 276
Db 1294 RVVEEKDALISQSRGKAFTQOIIEELKRLKEETKAKNALAHALQSAHDCDLREQYE 1353
Qy 277 ER-----NSKLR-----QENMELAERLKCLIEQVELREHIDKVF 311
Db 1354 EEOEAKGELQALSKANSEVAQWRKYETDAIQTREELEBAKKLAQRLQDAEHV---- 1409
Qy 312 KHKDQQQVDAKLAQAQAEMLKEAEERHQEKDFLLKEAVESQRMCELMKQETHLKQQL 371
Db 1410 -----EAVNAKAS-----LEKTKQLQNEVEDMLDVERANAAACAALDKKQNFDKIL 1458
Qy 372 ALYTKPEFEPQNTLSKSEVFTFPKQEMKMTKKIKLEKETMYRSWESSNKAALLEMA 431
Db 1459 AEWKQYIETQAELEASQKESRSLSSTELFKM-----KNAYEESLDHLQTLKRENILQOEI 1514
Qy 432 EERTVDRKELEGVQVKIQRLKLCRALQTERNDLNKRVDLSA 474
Db 1515 SDLT--EQIABGGKA-IHELEKVKQIQEKKSELQAALAEBA 1554

RESULT 31

```

O9DGM5 O9DGM5 PRELIMINARY; PRT; 1944 AA.
AC O9DGM5, 2001 (TREMELrel. 16, Created)
DT 01-MAR-2001 (TREMELrel. 16, Last sequence update)
DT 01-OCT-2003 (TREMELrel. 25, Last annotation update)
DE Fast myosin heavy chain isoform 2.
OS Gallus gallus (Chicken).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Archosauria; Aves; Neognathae; Galliformes; Phasianinae;
OC Gallus.
OX NCBI_TaxID=9031;
RN [1]
RP SEQUENCE FROM N.A.
RA Zhang Q., Bandman E.;
RT "Seven skeletal myosin heavy chain genes (MyHC) are organized as a
multigene complex in the chicken genome.";
RL Submitted (MAY-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF272033; AAF99314.1; -.
DR PIR; S24348; S24348.
DR HSSP; P13538; 2MYS.
DR GO; GO:0016459; C:myosin; IEA.
DR GO; GO:0005524; F:ATP binding; IEA.
DR GO; GO:0003774; F:motor activity; IEA.
DR InterPro; IPR000048; IQ_region.
DR InterPro; IPR001609; Myosin_head.
DR InterPro; IPR004009; Myosin_N.
DR InterPro; IPR002928; Myosin_tail.
DR Pfam; PF00612; IQ; 1.
DR Pfam; PF00063; myosin_head; 1.
DR Pfam; PF02736; Myosin_N; 1.
DR Pfam; PF01576; Myosin_tail; 1.
DR PRINTS; PR00193; MYOSINHEAVY.
DR ProDom; PD000355; myosin_head; 1.
DR SMART; SM00015; IQ; 1.
DR SMART; SM00242; MYSC; 1.
DR PROSITE; PS50096; IQ; 1.
SQ SEQUENCE 1944 AA; 232211 MW; 1547438220072D68 CRC64;

Query Match 9.6%; Score 259; DB 13; Length 1944;
Best Local Similarity 22.6%; Pred. No. 0.00022;
Matches 118; Conservative 82; Mismatches 181; Indels 142; Gaps 16;

QY 33 SQAPRKEGAQARTAQSGALRDVSEELSQLIEDILSTYCVNNQGGPGGAGGEPAPPE 92
DB 1096 SQTQSKIEDEQALGMQ---LQKKIKELQARIEEL-----EEIEARTSRA 1138
QY 93 DAESRTYVARNGEPEPTPVYGEKPSKGDPTNTEIRQSDVGDHRRPQKKAKGL 152
DB 1139 KAEXRADLSRELEE-----ISRLEEAGGATATQI-----DMNKKREAFQKMRD 1185
QY 153 GKEITLMTQTLNTLSTPEEKLAALCKKY---AELEEHNSQKMKLQKQKQKQK 208
DB 1186 LEEATL-----CHEATAAARKKHADSTALGEQIDNLQVKQKLEKSEKMEI 1236
QY 209 DHLRGESHKAVIARSKLESLELREL-----QRNRSLEKBEVQAR----- 248
DB 1237 DDLASNMESVSKANLEKMCRTLEDQLSKSEEHORMINDLSTQARLQTESGEYS 1296
QY 249 -----EEEEKKEVTSH-FQVTLNDIQLQMOH 276
DB 1297 RQVEKDALTSQLSRGKQAFQTOQIEELKRLHEEIKAKNALAHSQARHDCDILLREYE 1356
QY 277 ER-----NSKLR-----QENMELARLKKLIEQYELREBEHIDKVF 311
DB 1357 EEQKAGLQALSKANSEVAQMTKYETDAIQTEELERAKKLAQLQDAEHHV--- 1412
QY 312 KHKDLOQOLVDKIQQAQEMKKEABRHQRKQFLLKAEVSRQKQKQKQKQK 371
DB 1413 -----EAVNAKACAS-----LEKTKQLQNEVEDIMIDVERANACARLDKQKQK 1461
QY 372 ALYTEKEEFQNTLSKSEVPTTEKQEMKTKKIKKLEKETMYRWRWSSNKALLEWA 431

DB 1462 AEWKQKYETQAELEASQKESRSLSTELFKM---KNAYEESLDHLETLKRNKNLQORI 1517
QY 432 BEKTVRDKLEGLQVYKIQLEKLCALQTERNDLAKRVQDLQA 474
DB 1518 SOLT--EQIABGKA-IHELEKVKKQIQEOKSEIQALAEARA 1557

RESULT 32
Q8JG72 Q8JG72 PRELIMINARY; PRT; 1943 AA.
AC Q8JG72;
DT 01-OCT-2002 (TREMELrel. 22, Created)
DT 01-OCT-2002 (TREMELrel. 22, Last sequence update)
DT 01-OCT-2003 (TREMELrel. 25, Last annotation update)
DE Fast myosin heavy chain HcII.
GN N101.
OS Gallus gallus (Chicken).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Archosauria; Aves; Neognathae; Galliformes; Phasianinae;
OC Gallus.
OX NCBI_TaxID=9031;
RN [1]
RP SEQUENCE FROM N.A.
RA Zhang Q., Bandman E.;
RT "Seven skeletal myosin heavy chain genes (MyHC) are organized as a
multigene complex in the chicken genome.";
RL Submitted (MAY-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; AY116217; AAM8909.1; -.
DR PIR; A20923; A20923.
DR GO; GO:0016459; C:myosin; IEA.
DR GO; GO:0005524; F:ATP binding; IEA.
DR GO; GO:0003774; F:motor activity; IEA.
DR InterPro; IPR000048; IQ_region.
DR InterPro; IPR001609; myosin_head.
DR InterPro; IPR004009; Myosin_N.
DR InterPro; IPR002928; Myosin_tail.
DR Pfam; PF00612; IQ; 2.
DR Pfam; PF00063; myosin_head; 1.
DR Pfam; PF02736; Myosin_N; 1.
DR Pfam; PF01576; Myosin_tail; 1.
DR PRINTS; PR00193; MYOSINHEAVY.
DR ProDom; PD000355; myosin_head; 1.
DR PROSITE; PS50096; IQ; 1.
SQ SEQUENCE 1943 AA; 223383 MW; 7F57220B1EC83561 CRC64;

Query Match 9.5%; Score 257; DB 13; Length 1943;
Best Local Similarity 21.9%; Pred. No. 0.00027;
Matches 118; Conservative 83; Mismatches 181; Indels 158; Gaps 17;

QY 33 SQAPRKEGAQARTAQSGALRDVSEELSQLIEDILSTYCVNNQGGPGGAGGEPAPPE 92
DB 1096 SQTQSKIEDEQALGMQ---LQKKIKELQARIEEL-----EEIEARTSRA 1138
QY 93 DAESRTYVARNGEPEPTPVYGEKPSKGDPTNTEIRQSDVGDHRRPQKKAKGL 152
DB 1139 KAEXRADLSRELEE-----ISRLEEAGGATATQI-----DMNKKREAFQKMRD 1185
QY 153 GKEITLMTQTLNTLSTPEEKLAALCKKY---AELEEHNSQKMKLQKQKQKQK 208
DB 1186 LEEATL-----CHEATAAARKKHADSTALGEQIDNLQVKQKLEKSEKMEI 1236
QY 209 DHLRGESHKAVIARSKLESLELREL-----QRNRSLEKBEVQAR----- 248
DB 1237 DDLASNMESVSKANLEKMCRTLEDQLSKSEEHORMINDLSTQARLQTESGEYA 1296
QY 249 -----EEEEKKEVTSH-FQVTLNDIQLQMOH 276
DB 1297 RQADKKGGLISQLSRGKQAFQTOQIEELKRLHEEIKAKNALAHSQARHDCDILLREYE 1356
QY 277 ER-----NSKLR-----QENMELARLKKLIEQYELREBEHIDKVF 311
DB 1357 EELEKAGLQALSKANSEVAQMTKYETDAIQTEELERAKKLAQLQDAEHHV--- 1413

129 IRQDEVDRDRRPOEKAKGLGKBITLLMQLTNTLSTPEKLAALCKKYAELLBEHR 108
 76 ERPAZI--KPPKPEPTAPSNLEAL-----ERKAKENKILRELDWR 119
 189 NSQOMKLLQKQSQVORVDKHLRGHSHKAVLASKLESLELQHR---NRSLKBEVQ 245
 120 N-----KAKSAMGERDLRSEIKR---LKEELEKOEKELDKYIKISKOLKEX-LE 165
 246 RAREEEKRKRVTSHFOVTLNDIQLOMEOHNRNSKLR---QENKELAERLKLIRQY-E 301
 166 KAKRESEELKEKAEYERVEKI---AGKNELSKLELSDQNRRLAENLKKLKEKINE 222
 302 LREHIDKVPKHDLOQOLVDKILQQAQEMLEKEAERHOREKOPFLKAEVESQRMCELMK 361
 223 IKEER-DLKEETKEVGKLDQ-LQAKLQSLKEVKS---ERDDLANEVEALRNEKLR 276
 362 QOETHLKOQALYTEKEEFPQNTLSKSSEVFTTTPKOEKMTYKIKLKEETMYR---S 418
 277 KRIDKLASELNSLOKLDREKLEKARQHIGKLRBEIKRDRBEIKRKAQSKLDEIK 336
 419 RWESNKALL-----EMAEKTVDRKLEGLQVQKIORLEK 453
 337 RYEBGKLLVPPETEMAIK-VKGSIVIGKNSMVKALQE 374
 RESULT 35
 O08815 PRELIMINARY; PRT; 1206 AA.
 ID AC O08815;
 DT 01-JUL-1997 (TREMBlrel. 04, Created)
 DT 01-JUL-1997 (TREMBlrel. 04, Last sequence update)
 DT 01-OCT-2003 (TREMBlrel. 25, Last annotation update)
 DE Protein Kinase.
 GN SK2.
 OS Rattus norvegicus (Rat).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Rattus.
 OX NCBI_TaxID=10116;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=SD;
 RA Fukami Y., Yamamoto H., Ichihara T., Mori K., Gomi T., Sato K.;
 RT "SK2, a putative rat homologue of yeast protein kinase NRK1";
 RL Submitted (APR-1997) to the EMBL/GenBank/DBJ databases.
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=SD;
 RX MEDLINE=91287726; PubMed=2062320;
 RA Yue C.C.;
 RT "Novel putative protein kinase clones from a rat large granular
 lymphocyte tumor cell line";
 RL Mol. Immunol. 28:399-408(1991).
 CC -1. SIMILARITY: BELONGS TO THE SER/THR FAMILY OF PROTEIN KINASES.
 DR EMBL; AB003357; BAA20077.1; -.
 DR FIR; FT0204; FT0204.
 DR HSSP; P24941; 1CKP.
 DR GO; GO:0005524; F:ATP binding; IEA.
 DR GO; GO:0003677; F:DNA binding; IEA.
 DR GO; GO:0004518; F:nuclease activity; IEA.
 DR GO; GO:0004674; F:protein serine/threonine kinase activity; IEA.
 DR GO; GO:0016740; F:transferase activity; IEA.
 DR GO; GO:0006289; F:nucleotide-excision repair; IEA.
 DR GO; GO:0006468; P:protein amino acid phosphorylation; IEA.
 DR InterPro; IPR000719; Prot kinase.
 DR InterPro; IPR02290; Ser Thr kinase.
 DR InterPro; IPR008271; Ser Thr pkinase.
 DR InterPro; IPR001943; UvrE/C.
 DR Pfam; PF00069; pkinase; 1.
 DR SMART; PD000001; Prot kinase; 1.
 DR SMART; SM00220; S_TKc; 1.
 DR PROSITE; PS00107; "PROTEIN KINASE ATP; 1.
 DR PROSITE; PS50011; "PROTEIN_KINASE_DOM; 1.

DR PROSITE; PS00108; "PROTEIN_KINASE_ST; 1.
 DR PROSITE; PS50151; UVR; 1.
 KW ATP-binding; Kinase; Serine/threonine-protein kinase; Transferase.
 SQ SEQUENCE 1206 AA; 137888 MW; 8P5C6A19A3501PB9 CRC64;
 Query Match 9.4%; Score 253; DB 11; Length 1206;
 Best Local Similarity 22.2%; Pred. No. 0.00026;
 Matches 133; Conservative 92; Mismatches 199; Indels 176; Gaps 25;
 QY 12 EGQREPSQAAVAVERAGPGSSQA-----PR 37
 DB 682 ENAQELFVKAEP-----QAPASQASEPPVLPISINIHSENTENKEMGALPKPETILPP 737
 QY 38 KPEGAQARTAQSG--ALADVSEELSRQLEDILSTYCVDNNGCGPCDGAQEPAPEDA 94
 DB 738 EPENGKNDTSGTGSTVENSDDLNLSSFLS-----KTKDGSVSLQETRRQKTL 791
 QY 95 EKSRTYVARGPEPTPVVYGEKPSKGPNTETIRQSDVGDHRRRDEKKKAGLCK 154
 DB 792 KXTRKFFIDGVSVTVT---SKIVTDSKTEELR-----FLRQELRELELQK 838
 QY 155 EITLLMQTLN-TLSTPEEKLAALCKKYAELLBEHNSOKMKLLQKQQLVQEKDHLRG 213
 DB 839 EEQKAAQQLNGKLQOOREQIFRPFQ--EMLSKKQYDOEIEENLEKQKQTIE----- 889
 QY 214 EHSKAVLARSKLESICRELQRNRSIKKEGVORAREEKEKKEVTSHFOVTLNDIQLOME 273
 DB 890 -----RLE-----QETNRLRDE-AKRIKGEQKE---LSKFQNLRLN----- 923
 QY 274 QHNERNSKLRQSNMELAEKLLISQ-----YELREEH 306
 DB 924 RKKEGEFVKQKQQLDGLGALKKIIQQKAKLANIERECLNNTQQLLRAREAAIWELREH 983
 QY 307 IDKVPKHDLOQOLVDKILQQAQEMLEKAB---ERHOREKDFLLKEAV--ESQRMCELMK 361
 DB 984 LQE--KHOLLKQQLMDQVFIQRHLLKHEKETEQMYRNQRLIBELKNRQTQERARLPK 1041
 QY 362 QOETHLKOQALY-----TEKEEFPQNTLSKSSEVFTTTPKOEKMTYKIK 406
 DB 1042 IQRSKAKTNMFKKSLRINSTATPDQREKIKQFAAQBEK-----RQNERMAQH- 1092
 QY 407 KKLKETTMYRSRWBSSNKALILEMAEK-----TVDRKELE---GLQVKIORLEKLC 455
 DB 1093 QKHESQMRDLQOCRAVRELHLQNEKCHLIVHEHTQKLDELSEHSQELKEWR-EXLR 1151
 QY 456 RALQTERNDLNRVQDLS-----AGQGSLLTDSGPERPEGPGGAQAPSPRTEAPCYPG 510
 DB 1152 PRKKTLEEFARFKLQEQVFPFMTGESECLNPSAQSR-----GCLQTHSPSTRAPAWAG 1206
 RESULT 36
 O8AY28 PRELIMINARY; PRT; 1940 AA.
 ID AC O8AY28;
 DT 01-MAR-2003 (TREMBlrel. 23, Created)
 DT 01-MAR-2003 (TREMBlrel. 23, Last sequence update)
 DT 01-OCT-2003 (TREMBlrel. 25, Last annotation update)
 DE Fast myosin heavy chain HCIII.
 GN N127.
 OS Gallus gallus (Chicken).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;
 OC Gallus.
 OX NCBI_TaxID=9031;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Zhang Q.J., Bandman E.;
 RT "Seven Skeletal Myosin Heavy Chain Genes (MyHC) are Organized as a
 Multigene Complex in the Chicken Genome";
 RL Submitted (MAY-2002) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AY116218; AAM88910.1; -.
 DR GO; GO:0016459; C:myosin; IEA.
 DR GO; GO:0005524; F:ATP binding; IEA.

[illegible]

QY 399 MERMWKIKKLEKETTMYRSRWSSNKKALLEMAEKTVDKLEGLQVKIQRLKCRAL 458
DB 1155 VEHTQKLEDEHSHQELKWR-----EKLPRKTKLE-----EFPARKL 1195
QY 459 QTE 461
DB 1196 QEQ 1198
RESULT 38
Q07569 PRELIMINARY; PRT; 2139 AA.
AC Q07569; Q02504;
DT 01-NOV-1996 (T-EMBLrel. 01, Created)
DT 01-JUL-1997 (T-EMBLrel. 04, Last sequence update)
DT 01-OCT-2003 (T-EMBLrel. 25, Last annotation update)
DE Myosin heavy chain.
GN MYCA.
OS Entamoeba histolytica.
OC Eukaryota; Entamoebidae; Entamoeba.
CX NCBI_TaxID=5759;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=HMI:IMSS;
RX MEDLINE=93295430; PubMed=8515774;
RA Raymond-Denise A., Sansonetti P., Guillen N.;
RT "Identification and characterization of a myosin heavy chain gene
RT (mhcA) from the human parasitic pathogen Entamoeba histolytica.";
RL Mol. Biochem. Parasitol. 59:123-131(1993).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=HMI:IMSS;
RA Guillen N.;
RL Submitted (FEB-1997) to the EMBL/GenBank/DBJ databases.
DR EMBL; L03534; AAB48065.1; -;
DR PIR; T18296; T18296.
DR HSSP; P08799; IWND.
DR GO; GO:0016459; C:myosin; IEA.
DR GO; GO:0005524; F:ATP binding; IEA.
DR GO; GO:0003774; F:motor activity; IEA.
DR InterPro; IPR000048; IQ region.
DR InterPro; IPR001609; myosin_head.
DR Pfam; PF00612; IQ; 2.
DR Pfam; PF00063; myosin_head; 1.
DR PRINTS; PR00193; MYOSINHEAVY.
DR ProDom; PD000355; myosin_head; 1.
DR SMART; SM00242; MYSC; 1.
SQ SEQUENCE 2139 AA; 245225 MW; C68307341DB51DD1 CRC64;
Query Match 9.2%; Score 249.5; DB 5; Length 2139;
Best Local Similarity 21.7%; Pred. No. 0.00057;
Matches 106; Conservative 93; Mismatches 172; Indels 117; Gaps 19;
QY 54 DVSELSRQLDILSTYC-VNNMGQGGEDGAQGE-----PAPPE 92
DB 1169 DVSOYLOKQKEYESQIAKMQEKEAIGNDVYKKEITIKKELSIQSLQKLDTEYEKE 1228
QY 93 DAESRTVVRNGEPPTFPVYVGEKSPGDPNTEIRQSDVGDGRHRRPOEKKKAGL 152
DB 1229 DAERKKKEIEKE-----MKALQKEKE-----NVESKKNSTE-----KDKKKLEDN 1268
QY 153 KEETLLMQTLNTISTPEKLAALCK-----KYALLEHRRNSQOMKLLQKQSLVQEK 208
DB 1269 LKD---TQKLLDMDTADNKKLAKAKDLQALNEVQNHKAVADAEILNKKKQAQSKEL 1325
QY 209 DHLRGEHSAVLARSLKSLCRLQHRNRLSKESGVOPAREEKEKKEVTSHFQVTLNDI 268
DB 1326 NSLKAEALTKAKSVESKNSKNGENKAAUSEE-IDQANEK-----LKNIQADLRKA 1377
QY 269 QLQMEHNRNLSKQENNELA---ERLKLIEQVLEAE-----EHIDKVFHKDLQ 318
DB 1378 TADLQEAENEKAEVBAQRDLVADNKKTKLEIKARDEENTYKVENYKVLKREK---- 1433

QY 319 QLVDAKLOQAQEMLK-BAEERHQREKDFLLKEAVESQRMCELMKQOETHLQKQOLA----L 373
DB 1434 ---EADLEENENLDIEKKQRMKEQV---KLEGE-----LKETDKLAAIAEKDSI 1482
QY 374 YTEK-----PEEFQNTLSKSEVPTTFQKQEMKWTKKIKLEKETTYRSWSSNKKAL 427
DB 1483 PTAKKQSDADLEELNKTVEEHDEVVAKLNTQITKLTRDNQSAABELNELRSKADKKKKI 1542
QY 428 LEMAE-----EKTVD-----KELEGVQVKIQRLKCRALQTER 462
DB 1543 SELEQVNELSRPPVGTGADENELKIRDAQLADLNKALEMKGVQNNQQAATNKELKAKD 1602
QY 463 NDLNKRVQ 470
DB 1603 NDLTSKIE 1610
RESULT 39
Q9CS72 PRELIMINARY; PRT; 610 AA.
AC Q9CS72;
DT 01-JUN-2001 (T-EMBLrel. 17, Created)
DT 01-JUN-2001 (T-EMBLrel. 17, Last sequence update)
DT 01-OCT-2002 (T-EMBLrel. 22, Last annotation update)
DE 5730485H2IRik protein (Fragment).
GN 5730485H2IRIK.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Euthera; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
CX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=Embryo;
RX MEDLINE=21085860; PubMed=11217851;
RA Kawai J., Shinagawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,
RA Arakawa T., Hara A., Fukunishi Y., Konno H., Adachi J., Fukuda S.,
RA Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamanaka I.,
RA Saito T., Okazaki Y., Gotohori T., Bono H., Kasukawa T., Saito R.,
RA Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,
RA Fleischmann W., Gaasterland T., Gissi C., King B., Kochiwa H.,
RA Kuehl P., Lewis S., Matsuo Y., Nikaido I., Pesole G., Quackenbush J.,
RA Schirml L.M., Staubli P., Suzuki R., Tomata M., Wagner L., Washio T.,
RA Sakai K., Okido T., Furuno M., Aono H., Baldarelli K., Barsh G.,
RA Blake J., Boffelli D., Bojunga N., Carninci P., de Bonaldo M.F.,
RA Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M.,
RA Gustincich S., Hill D., Hofmann M., Hume D.A., Kamiya M., Lee N.H.,
RA Lyons P., Marchionni L., Mashima J., Mazzarelli J., Mombarts P.,
RA Nordone P., Ring B., Ringwald W., Rodriguez I., Sakamoto N.,
RA Sasaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-P.,
RA Suzuki H., Toyooka K., Wang K.H., Weitz C., Whittaker C., Wilming L.,
RA Wynshaw-Boris A., Yoshida K., Hasegawa Y., Kawaji H., Kohetsuki S.,
RA Hayashizaki Y.;
RT "Functional annotation of a full-length mouse cDNA collection.";
RL Nature 409:585-590(2001).
DR EMBL; AK017709; BAB30888.1; -;
DR MGD; MGI:1917848; 5730485H2IRik.
DR InterPro; IPR000533; Tropomyosin.
DR PRINTS; PR00194; TROPOMYOSIN.
FT NON TER 610 610
SQ SEQUENCE 610 AA; 70346 MW; 9B9EFC87F64FCE80 CRC64;
Query Match 9.2%; Score 249; DB 11; Length 610;
Best Local Similarity 24.3%; Pred. No. 0.0002;
Matches 129; Conservative 86; Mismatches 217; Indels 98; Gaps 21;
QY 10 GP-EGQERPSQAAPAVEAEGPGSQAPRKPEGAQARTAQSGALRDVSELSRQLEDILS 68
DB 31 GPSADAKKANKRKGEDDVMASGTVRHLKPSGESEKTKKPL-----ELSK--EDLIQ 82
QY 69 TYCV-----DNNQGGPDGAGQGEPAEPEDAEKRTYVARGEPPTVPVYGEKE 118
DB 83 LLSIMEGELQVGRMSRNOQ--GSSSNHGVSCPK-----SIISDGGKGPSEDAKKKA 135

```
QY 119 PSKGDNTBIROSDEVDGRDRHP-----QEKKAKGLCKE--ITLMTLNTLSTPE 171
Db 136 NRKEG---DDVMASGV--XHLKPSGESEKTKKPLESKEDLIQLLSIMSELOARD 190
QY 172 KLAALC--KKYABELBEHNSQOMKLLQ--KKOSOLVQEKDHLRGEHSAVLARSKLS 227
Db 191 VIHMLATEKTPVLEAHYGSAPPEKVLRLVHRDAILAQEKSTIGEDVVEKPISELDRLE 250
QY 228 LCELOL---HNSLKEEGVQARAREEKEKEVTSHPQVTLNDIQLQOQHNRNSKLRQ 284
Db 251 KQKETVRMLLEQLLAKCHRRTVYLENEKHKHTDYMKSDFTNLLQERLERLKLLE 310
QY 285 E-----NMELAEKLLKLEQ-----YELREHIDKVPKHDLOQLVD--- 322
Db 311 QEKAYOARKEKENAKLNLKRLDELVKLSPALMLVDERQWHIEQL-----GLSQKVQDLT 366
QY 323 AKLQOAEMLKAEERHOREKDFLLKEAV-----ESQRCMLKMQQETHLQO-- 369
Db 367 QKURSEEEKLKALTYKSKDRQKLLKLEVDPEKASRFQOEHEENAKLANOESHNRQIR 426
QY 370 ----QLALYTEKPEFTQNTLSKSEVFTTFKQEMERTKKIKLEKETTYMSRWESSNK 425
Db 427 LKLVGLSQRIEELEETNKSLOKAE-----ELQELRDKIAKGECSNLSMAEVESLAK 479
QY 426 ALLEMA---BEKTVRKLEGLQVKIQRLEKLCRALQTERNDLNKRVQDL 472
Db 480 RVLEMEGKDEITKTEAQCRELKKLQBEHHSKELRLEVEKLQKRMSEL 529

RESULT 40
Q91LX9 PRELIMINARY; PRT; 1003 AA.
AC Q91LX9
DT 01-DEC-2001 (TReMBLrel. 19, Created)
DT 01-DEC-2001 (TReMBLrel. 19, Last sequence update)
DT 01-OCT-2003 (TReMBLrel. 25, Last annotation update)
DE ORF73.
OS Kaposi's sarcoma-associated herpesvirus (KSHV) (Human herpesvirus 8).
OC Viruses; dsDNA viruses, no RNA stage; Herpesviridae;
OC Gammaherpesvirinae; Rhadinovirus.
OX NCBI_TaxID=37296;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=21376412; PubMed=11483733;
RA Garber A.C., Shu M.A., Hu J., Renne R.;
RT "dna binding and modulation of gene expression by the latency-
RT associated nuclear antigen of Kaposi's sarcoma-associated
RT herpesvirus.";
RL J. Virol. 75:7882-7892(2001).
DR EMBL: AF360120; AAK50002.1; -.
DR InterPro: IPR002017; Spectrin.
DR InterPro: IPR000533; Tropomyosin.
DR PRINTS: PR00194; TROPOMYSIN.
SQ SEQUENCE 1003 AA; 115517 MW; C20C43308B01A0A3 CRC64;

Query Match 9.2%; Score 249; DB 12; Length 1003;
Best local similarity 20.0%; Pred. No. 0.00033;
Matches 103; Conservative 108; Mismatches 217; Indels 86; Gaps 13;

QY 11 PEGAQERPSQAAPVAEAGGSGQAQRPK-----EGAAQ-RTAQSGALRDVS 56
Db 236 PPMATESPIYVGSSDGDTP-----PRQPTSPISIGSSPSGSGMGDDTAMLVLLAEIA 290
QY 57 HELSRQLEDILTSCYVDNNOGG-----PGEDCAQCEPAPPEDAKS 97
Db 291 HEASKNEKE-----CSENKQAGEDGDNLSKESQVYKDDNKNKDEEEQETDEDEDD 345
QY 98 RTTVARNGE-----PETPVVYGEKPSKGDPNTEIRQSGVDGRDRHP 143
Db 346 EEDDEEDDEEDDEEDDEEDDEEDDEEDDEEDDEEDDEEDDEEDDEEDDEEDDEEDDEED 405
QY 144 QEKKAK---GLGKETITLMTLNTLSTPEKLAALCKKYALLBEHNSQOMKLLQK 199
```

```
Db 406 BEDKKEDEEDGDGNK-TLSIQSSQQQEPQO-----QBPOQEPQOQBPOQEPLOE 457
QY 200 KQSOLVQEKDHLRGEHSAVLARSKLSLCELORHNRSLKEEGVQARAREEKEKSVTS 259
Db 458 PQQQEPQOQBPOQEPQO-----BPQOQEPQOQBPOQEPQOQBPOQEPQOQBPOQ 510
QY 250 HPQVTLNDIQLQME-QHNERNKLNQENNELAEKLLKLEIQVELEEEHIDKVPKHDLOQ 318
Db 511 QDEQEQBDEQDEQDEQDEQDEQDEQDEQDEQDEQDEQDEQDEQDEQDEQDEQDEQDE 569
QY 319 QLVDAKLAQQAQMLKAEERHOREKDFLLKEAVSQRCHLMKQETHLQQLALYTEKF 378
Db 570 QELREHIDKVPKHDLOQQLALYTEKF-----ELREHIDKVPKHDLOQQLALYTEKF 621
QY 379 REFQNTLSKSEVFTTFKQEMERTKKIKLEKETTYMSRWESSNKALLENABEKTVRD 438
Db 622 EBOEELREHIDKVPKHDLOQQLALYTEKF-----ELREHIDKVPKHDLOQQLALYTEKF 681
QY 439 KELEGQVKIQRLEKLCRALQTERNDLNKRVQDL 472
Db 682 QELREHIDKVPKHDLOQQLALYTEKF-----ELREHIDKVPKHDLOQQLALYTEKF 715

RESULT 41
Q9P216 PRELIMINARY; PRT; 1005 AA.
AC Q9P216; Q96L75;
DT 01-OCT-2000 (TReMBLrel. 15, Created)
DT 01-OCT-2000 (TReMBLrel. 15, Last sequence update)
DT 01-OCT-2003 (TReMBLrel. 25, Last annotation update)
DE Hypothetical protein KIAA1361 (Serine/threonine kinase TA01)
DE (fragment).
GN KIAA1361.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Brain;
RX MEDLINE=20181126; PubMed=10718198;
RA Nagase T., Kikuno R., Ishikawa K., Hirose M., Ohara O.;
RT "Prediction of the coding sequences of unidentified human genes XVI.
RT The complete sequences of 150 new cDNA clones from brain which code
RT for large proteins in vitro.";
RL DNA Res. 7:65-73(2000).
RN [2]
RP SEQUENCE OF 5-1005 FROM N.A.
RA Jenkins S.G., D'Andrea R.J., Gamble J.R., Vadas M.A.;
RT "Characterization of human TA01.";
RL Submitted (JUL-2001) to the EMBL/GenBank/DBJ databases.
CC -1- SIMILARITY: BELONGS TO THE SER/THR FAMILY OF PROTEIN KINASES.
DR EMBL: AB037782; BA032599.1; -.
DR EMBL: AY049015; AAL12217.1; -.
DR HSP; P24941; I338.
DR GO: GO:0005524; P:ATP binding; IEA.
DR GO: GO:0004674; P:protein serine/threonine kinase activity; IEA.
DR GO: GO:0004674; P:transferase activity; IEA.
DR GO: GO:0006468; P:protein amino acid phosphorylation; IEA.
DR InterPro: IPR000719; Prot kinase.
DR InterPro: IPR002290; Ser_thr_kinase.
DR InterPro: IPR008271; Ser_thr_kinase.
DR Pfam: PF00069; pkinase; I.
DR ProDom: PD000001; Prot kinase; 1.
DR SMART: SM00220; S_TKc_1.
DR PROSITE: PS00107; PROTEIN KINASE ATP; 1.
DR PROSITE: PS00111; PROTEIN KINASE DOM; 1.
DR PROSITE: PS00108; PROTEIN KINASE ST; 1.
KW Hypothetical protein; ATP-binding; Kinase;
KW Serine/threonine-protein kinase; Transferase.
FT NON_TER 1
SQ SEQUENCE 1005 AA; 116540 MW; A83BD4E53569BCB5 CRC64;
```

```

Query Match      9.2%; Score 248.5; DB 4; Length 1005;
Best Local Similarity 21.6%; Pred. No. 0.00035;
Matches 139; Conservative 117; Mismatches 197; Indels 191; Gaps 29;

QY 23 PAVEAE-----GPGSSQA--PRKPEGAQARTAGSALRDVSELSRQLE 64
DB 328 PAVEAQEEEEEDHGVRGTGTVNSVGSNQSIIPMSISASSQSSVNSLPDVSDDKS--ELD 386
QY 65 DILSTYCVNNQ-----GPGEDG--AQGEPAEPEDAESRTTVARNGEPEPTPVVYGEKE 118
DB 387 MMEGDHTVMSNSVHLKPEENYQ--EGEDPR-----TRASAPQSPQVSRHSHY 434
QY 119 PSKGDPTNEERQSDVEGDRDRHRRPOEKKAKGLGKEIT--LIMQTLNTLSTPEEKLA 174
DB 435 HYNRNHPATIRTSALV---TRQMEHQDSSELREQSGYKMRQRQKQMLTLENKLA 490
QY 175 ALCKYA---ELIEHRN---SOKMKLLQKQSOVLQV-----KDLRGHSHK 217
DB 491 AEMDEHLRLDKDLETQRNFAAEKELIKKHQASMEKAKVMANEKKFOQHQAQOKK 550
QY 218 AVLARSKLESCLRELQRHNRSLKEGVOR---AREBEE---KRKEVTSHPQV--TLNDI 268
DB 551 EL--NSFLSQREYKRLKEQLKEELNENQSTPKKEQEWLSKQKENIQHFQABEEANLL 608
QY 269 QLOWE-----QHN-----ERNKLRQENMELAEKRLKLIQOYELRE 304
DB 609 RRQRYLELCRRFRKRRMLLGRHNLQDLVREELNKRQTQDKLEHAWMLLRHSHESMQBLEF 668
QY 305 EHIDKVFHKDLQOOLVDKLAQQAQEMLEKAEERHOREKDFLLKEAVESQRMCELMKQOE 364
DB 669 RLNTIQK---MRCELI---RLQHQTELNTQLEYKRRERELRRKHVMEVQOPKSLKSE 723
QY 365 THLQOOL---ALYTEKPEEFQNTLSKSSSVFTTFKQEMEKTKIKKLEKETTMYRSRW 420
DB 724 LQIKKQFDQTKIQTRQYKALRNHLE---TTPKSEHKAVLKRLK--BEQTRKLAIAE 776
QY 421 ESSKALLEMAEKTIVR---QHN-----ERNKLRQENMELAEKRLKLIQOYELRE 442
DB 777 EYDHSINEMLSQALRLDRAQEAECQVLKMLQOOLELLNAYOSKIKMQABEAQHRELRL 836
QY 443 GLOVKI-----QRLEKLCRALQTERND-----LNKRVDLSAGQGS-----L 480
DB 837 ELEQVSLRRALLQKIEEEMALQNERTERIRSLERQAREIEAFDSERMLGFSNMVL 896
QY 481 TDSGPRPEPGQAQAPSSPVTRAPCYGAP--STASQGTGP 522
DB 897 SNLSPE-----AFSHSYPGASGWSHNPTGCGPG 924

RESULT 42
OB8664 PRELIMINARY; PRT; 1001 AA.
AC OB8664;
DT 01-NOV-1998 (Tremblrel. 08, Created)
DT 01-NOV-1998 (Tremblrel. 08, Last sequence update)
DT 01-OCT-2003 (Tremblrel. 25, Last annotation update)
DE Serine/threonine protein kinase TA01.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=9903202; PubMed=9786855;
RA Hutchison M., Berman K.S., Cobb M.H.;
RT "Isolation of TA01, a protein kinase that activates MEKs in stress-
RT activated protein kinase cascades."
RL J. Biol. Chem. 273:28625-28632(1998).
CC -!- SIMILARITY: BELONGS TO THE SER/THR FAMILY OF PROTEIN KINASES.
DR EMBL; AF084205; AAC71014.1; --
DR PIR; T17365; T17365.
DR HSP; P24941; 1B38.
DR GO; GO:0005524; F:ATP binding; IEA.
GO; GO:0004674; F:protein serine/threonine kinase activity; IEA.
GO; GO:0016740; F:transferase activity; IEA.
GO; GO:0006468; F:protein amino acid phosphorylation; IEA.
DR InterPro; IPR00719; Prot kinase.
DR InterPro; IPR002290; Ser thr pkinase.
DR InterPro; IPR008271; Ser thr pkin_AS.
DR Pfam; PF00069; pkinase; 1.
DR ProDom; PD000001; Prot kinase; 1.
DR SMART; SM00220; S_TKc_1.
DR PROSITE; PS00107; PROTEIN KINASE_ATP; 1.
DR PROSITE; PS00111; PROTEIN KINASE_DOM; 1.
DR PROSITE; PS00108; PROTEIN KINASE_ST; 1.
KW ATP-binding; Kinase; Serine/threonine-protein kinase; Transferase.
SQ SEQUENCE 1001 AA; 115952 MW; 85511B62DBD62FCC CRC64;

Query Match      9.1%; Score 247; DB 11; Length 1001;
Best Local Similarity 21.3%; Pred. No. 0.00041;
Matches 142; Conservative 118; Mismatches 212; Indels 176; Gaps 30;

QY 23 PAVEAE-----GPGSSQA--PRKPEGAQARTAGSALRDVSELSRQLE 64
DB 324 PAVEAQEEEEEDHGVRGTGTVNSVGSNQSIIPMSISASSQSSVNSLPDASDDKS--ELD 382
QY 65 DILSTYCVNNQ-----GPGEDG--AQGEPAEPEDAESRTTVARNGEPEPTPVVYGEKPS 120
DB 383 MMEGDHTVMSNSVHLKPEENYQ--EGEDPR-----TRASAPQSPQVSRHSHY 432
QY 121 KGDPTNEERQSDVEGDRDRHRRPOEKKAKGLGKEIT--LIMQTLNTLSTPEEKLAAL 176
DB 433 RNREHPATIRTSALV---TRQMEHQDSSELREQSGYKMRQRQKQMLTLENKLAKE 488
QY 177 CKYA---ELIEHRN---SOKMKLLQKQSOVLQV-----KDLRGHSHKAV 219
DB 489 MDEHLRLDKDLETQRNFAAEKELIKKHQASMEKAKVMANEKKFOQHQAQOKKEL 548
QY 220 LARSKLESCLRELQRHNRSLKEGVOR---AREBEE---KRKEVTSHPQV--TLNDIQL 270
DB 549 --NSFLSQREYKRLKEQLKEELNENQSTPKKEQEWLSKQKENIQHFQABEEANLLRR 606
QY 271 QME-----QHN-----ERNKLRQENMELAEKRLKLIQOYELREEH 306
DB 607 QRQYLELCRRFRKRRMLLGRHNLQDLVREELNKRQTQDKLEHAWMLLRHSHESMQBLEFRH 666
QY 307 IDKVFHKDLQOOLVDKLAQQAQEMLEKAEERHOREKDFLLKEAVESQRMCELMKQOETH 366
DB 667 INTIQK---MRCELI---RLQHQTELNTQLEYKRRERELRRKHVMEVQOPKSLKSELO 721
QY 367 LKQOL---ALYTEKPEEFQNTLSKSSSVFTTFKQEMEKTKIKKLEKETTMYRSRWES 422
DB 722 IKQPDQDQTKIQTRQYKALRNHLE---TTPKSEHKAVLKRLK--BEQTRKLAIAEQ 774
QY 423 SNKALLEMAEKTIVR---QHN-----ERNKLRQENMELAEKRLKLIQOYELRE 444
DB 775 YDHSINEMLSQALRLDRAQEAECQVLKMLQOOLELLNAYOSKIKMQABEAQHRELREL 834
QY 445 QVKI-----QRLEKLCRALQTERND-----LNKRVDLSAGQGS-----LTD 482
DB 835 EQRVSLRRALLQKIEEEMALQNERTERIRSLERQAREIEAFDSERMLGFSNMVLN 894
QY 483 SGPRPEPG--PGAQAP--SPVTRAP--CYGAPSTEASG---QTGQP 523
DB 895 LSPEAFSHSYPGASGWSHNPTGSGPHWGHMGTPQAWGHPWQCGPG 942

RESULT 43
OB85C7 PRELIMINARY; PRT; 2055 AA.
AC OB85C7;
DT 01-JUN-2002 (Tremblrel. 21, Created)
DT 01-JUN-2002 (Tremblrel. 21, Last sequence update)
DT 01-OCT-2003 (Tremblrel. 25, Last annotation update)
DE Chimeric erythrocyte-binding protein MAEBL.
GN MAEBL.

```

```
OS Plasmodium falciparum.
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX NCBI_TaxID=5833;
RN [1]
RP SEQUENCE FROM N.A.
RA Blair B.L., Kappe S.H.I., Maciel J.B., Balu D.J., Adams B.;
RT "Plasmodium falciparum maeli is a unique member of the ebl family.";
RL Mol. Biochem. Parasitol. 0:0-0(2002).
RN [2]
RP SEQUENCE FROM N.A.
RA Michon P., Stevens J.R., Kaneko O., Adams J.H.;
RT "Plasmodium falciparum maeli is a unique member of the EBL family.";
RL Mol. Biol. Evol. 0:0-0(2002).
RN [3]
RP SEQUENCE FROM N.A.
RA Chal M., Dutta S., Ockenhouse C.F.;
RT "Identification, expression and functional characterization of MAEBL,
RT a sporozoite and asexual blood stage chimeric erythrocyte binding
RT protein of Plasmodium falciparum.";
RL Mol. Biochem. Parasitol. 0:0-0(2002).
DR EMBL; AF042084; AAL10509.1; -.
DR EMBL; AF400002; AAM90625.1; -.
DR HSSP; P04288; IIC2.
DR InterPro; IPR008602; Duffy binding.
DR Pfam; PF05424; Duffy binding; 1.
SQ SEQUENCE 2055 AA; 243210 MW; 783P48BB7286B1F3 CRC64;

Query Match          9.1%; Score 246.5; DB 5; Length 2055;
Best Local Similarity 24.5%; Pred. No. 0.0009;
Matches 101; Conservative 85; Mismatches 146; Indels 81; Gaps 16;

QY 116 EKPSKGDPTTEIRQSDV-----GDRHRRPQEKKAAGLKGKEITLLMTLNTL 166
DB 1436 EKKAENLKAAEKKKADLKKAEKKKADLKKAEKKKADLKKAE--EKKADEL 1494
QY 167 STPEKLAALCKKYAEL--LEHRNSOKQMLKQKQOLVQEKDHLRGEHSKAVLARSK 224
DB 1495 KKAEEK-----KKADELKKAEEKKADLKKAEKKKADLKKAEELKKAEEK-----KK 1544
QY 225 LESLRELQRHRSIKE-----EGVQARAEEREEKKKEVTSHFQVTLNDIQLQM 272
DB 1545 VEQKKEERERNMALRAAILKQIEKKRIEVMKLYEEKKKAE-----QLKK 1593
QY 273 EQHNRNSKLQENMELAEKLKLIQYELREHIDKVFHKDLQOOLVDKLAQQAQEML 332
DB 1594 EE-----EEKIKAEQAKKEEEKKKVQOLKKKEEEKK--KAEQKKKEEENKIK--AEQL 1645
QY 333 KEAEERHOREKDFLLKAEVDSQRMCELMKQOETHLK--QOLALYTEKPEFQNTLSKSE 390
DB 1646 KKEEEKKKAEELKKEEEKKKAEQAKKEEEKKKVQOLKKKEEEKKKAEQAKKEE 1705
QY 391 ----VFTTFQEMERTKIKKLEKETTMYRSRWESSNKALLEMAEKTVDKLEGLQV 446
DB 1706 ENKIKVQOLKKEEEKKKAEELKKE-----EEKKKVQOLKKEE---EKKAEEIRK 1754
QY 447 KIQLR--EKLCLALQTERNDLNRVQDLSAGGSGSLTDSGPERRPGGAQAP 497
DB 1755 EKEAVIEELKKEDKKRMEVEKKIKD-----TKDNFENIQEGNNKNTP 1798

RESULT 44
Q8IHP3
ID Q8IHP3 PRELIMINARY; PRT; 2055 AA.
AC Q8IHP3;
DT 01-MAR-2003 (TrEMBLrel. 23, Created)
DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE MAEBL, putative.
GN Pf11_0486.
OS Plasmodium falciparum (isolate 3D7).
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX NCBI_TaxID=36329;
RN [1]
```

```
RP SEQUENCE FROM N.A.
RC STRAIN=3D7;
RX MEDLINE=22255705; PubMed=12369864;
RA Gardner M.J., Hall N., Fung B., White O., Berriman M., Hyman R.W.,
RA Carlton J.M., Pain A., Nelson K.E., Bowman S., Paulsen I.T., James K.,
RA Eisen J.A., Rutherford K., Salzberg S.L., Craig A., Kyes S.,
RA Chan M.-S., Renev V., Shallow J.J., Suh B., Peterson J., Anguoli S.,
RA Petrea M., Allen J., Seilong J., Haft D., Mather M.W., Vaidya A.S.,
RA Martin D.M.A., Fairlamb A.H., Fraunholz M.J., Roos D.S., Ralph S.A.,
RA McFadden G.I., Cummings L.M., Subramanian G.M., Mungall C.,
RA Venter J.C., Carucci D.J., Hoffman S.L., Newbold C., Davis R.W.,
RA Fraser C.M., Barrell B.;
RT "Genome sequence of the human malaria parasite Plasmodium
RT falciparum."
RL Nature 415:498-511(2002).
DR EMBL; AB014843; AAN36066.1; -.
DR InterPro; IPR008602; Duffy binding.
DR Pfam; PF05424; Duffy binding; 1.
SQ SEQUENCE 2055 AA; 243210 MW; 783P48BB7286B1F3 CRC64;

Query Match          9.1%; Score 246.5; DB 5; Length 2055;
Best Local Similarity 24.5%; Pred. No. 0.0009;
Matches 101; Conservative 85; Mismatches 146; Indels 81; Gaps 16;

QY 116 EKPSKGDPTTEIRQSDV-----GDRHRRPQEKKAAGLKGKEITLLMTLNTL 166
DB 1436 EKKAENLKAAEKKKADLKKAEKKKADLKKAEKKKADLKKAE--EKKADEL 1494
QY 167 STPEKLAALCKKYAEL--LEHRNSOKQMLKQKQOLVQEKDHLRGEHSKAVLARSK 224
DB 1495 KKAEEK-----KKADELKKAEEKKADLKKAEKKKADLKKAEELKKAEEK-----KK 1544
QY 225 LESLRELQRHRSIKE-----EGVQARAEEREEKKKEVTSHFQVTLNDIQLQM 272
DB 1545 VEQKKEERERNMALRAAILKQIEKKRIEVMKLYEEKKKAE-----QLKK 1593
QY 273 EQHNRNSKLQENMELAEKLKLIQYELREHIDKVFHKDLQOOLVDKLAQQAQEML 332
DB 1594 EE-----EEKIKAEQAKKEEEKKKVQOLKKKEEEKK--KAEQKKKEEENKIK--AEQL 1645
QY 333 KEAEERHOREKDFLLKAEVDSQRMCELMKQOETHLK--QOLALYTEKPEFQNTLSKSE 390
DB 1646 KKEEEKKKAEELKKEEEKKKAEQAKKEEEKKKVQOLKKKEEEKKKAEQAKKEE 1705
QY 391 ----VFTTFQEMERTKIKKLEKETTMYRSRWESSNKALLEMAEKTVDKLEGLQV 446
DB 1706 ENKIKVQOLKKEEEKKKAEELKKE-----EEKKKVQOLKKEE---EKKAEEIRK 1754
QY 447 KIQLR--EKLCLALQTERNDLNRVQDLSAGGSGSLTDSGPERRPGGAQAP 497
DB 1755 EKEAVIEELKKEDKKRMEVEKKIKD-----TKDNFENIQEGNNKNTP 1798

RESULT 45
Q9WU41
ID Q9WU41 PRELIMINARY; PRT; 1202 AA.
AC Q9WU41;
DT 01-NOV-1999 (TrEMBLrel. 12, Created)
DT 01-NOV-1999 (TrEMBLrel. 12, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Ste20-related kinase SMAK.
GN STK2 OR SMAK.
OS Mus musculus (Mouse).
OC Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=BALB/c;
RA Sabourin L.A., Seale P., Tamai K., Wagner J., Rudnicki M.A.;
RT "Caspase 3 cleavage of the Ste20-related kinase SLK releases and
RT activates an apoptosis-inducing kinase domain and an actin-
RT disassembling region.";
```

```
RL  Mol. Cell. Biol. 20:2949-2949 (2000).
RN  [2]
RP  SEQUENCE FROM N.A.
RC  STRAIN=BALB/c;
RX  MEDLINE=20073281; PubMed=10611247;
RA  Sabourin L.A., Seale P., Wagner J., Rudnicki M.A.;
RT  "Caspase 3 cleavage of the Ste20-related kinase SLK releases and
XT  activates an apoptosis-inducing kinase domain and an actin-
XT  disassembling region.";
RL  Mol. Cell. Biol. 20:684-696(2000).
CC  -!- SIMILARITY: BELONGS TO THE SER/THR FAMILY OF PROTEIN KINASES.
DR  EMBL; AF112855; AAD28717.1; -.
DR  HSSP; P24941; LCKP.
DR  MGD; MGI:103241; Stk2.
DR  GO; GO:0005524; F:ATP binding; IEA.
DR  GO; GO:0003677; F:DNA binding; IEA.
DR  GO; GO:0004518; F:nuclease activity; IEA.
DR  GO; GO:0004674; F:protein serine/threonine kinase activity; IEA.
DR  GO; GO:0016740; F:transferase activity; IEA.
DR  GO; GO:0006289; P:nucleotide-excision repair; IEA.
DR  GO; GO:0006458; P:protein amino acid phosphorylation; IEA.
DR  InterPro; IPR000719; Prot_kinase.
DR  InterPro; IPR002290; Ser_thr_kinase.
DR  InterPro; IPR008271; Ser_thr_pkinase.
DR  InterPro; IPR001943; UvrB/C.
DR  Pfam; PF00069; pkinase; 1.
DR  ProDom; PD000001; Prot_kinase; 1.
DR  SMART; SM00220; S_TKC_1.
DR  PROSITE; PS00107; PROTEIN_KINASE_ATP; 1.
DR  PROSITE; PS00111; PROTEIN_KINASE_DOM; 1.
DR  PROSITE; PS00108; PROTEIN_KINASE_ST; 1.
DR  PROSITE; PS00151; UVR; 1.
KW  ATP-binding; Kinase; Serine/threonine-protein kinase; Transferase.
SQ  SEQUENCE 1202 AA; 137713 MW; 9F4C78A4C31FAAB2 CRC64;

Query Match          9.0%; Score 244; DB 11; Length 1202;
Best Local Similarity 21.2%; Pred. No. 0.00069;
Matches 133; Conservative 99; Mismatches 183; Indels 212; Gaps 28;

QY  6 QPAGPP--GAQERP-----SQAPAV-BAEG--PGSSQAPKPGGAQARTAQSG 50
DB  :|||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
DB  582 KPTGPGAGGAEPPGGVERVEDKQPEQQAQVCAEGQLTSTSTTR----ATLEQPEYD 637
QY  51 ALRDVSELS-ROLEDILTSTVVDNNOGGGPDG-----AQ 85
DB  :|||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
DB  638 EYEQVSESNSEIELERLVVTGAERALGSGEAAATEVDLERKENAQKVPVKAESQAPAA 697
QY  86 GEPAPPE-----DAKSRITVARNGEPEPTPVVYGEKPSKG---DPNT-----126
DB  :|||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
DB  698 SQSEPEFVLIPININSETTENKEEMGALPKPETILPPEPEHEKGNDDTSGTGSTVENS 757
QY  127 -----BEIROSDEVDGRDHRPQE--KKAKGL--GKITLLMQTLTLSTPE 170
DB  :|||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
DB  758 SGDLNLSSIFLSKAKDSGVSLSQETRRQKTLKTRKFIVDGVVSV---TTSKIIVDS 814
QY  171 EKLAALCKYAELEHRSNQKMKLLOKQ-----SOLVQEKDHLRGEHSKAVLAR 222
DB  :|||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
DB  815 D-----SKTEELFRLRQELSRLLQKEEQRAQOLNGLQOQREQIPRPEQEMLSK 868
QY  223 SK-----LESCLQRHNRSLKEEGVQARREE-----EKREKVTSHFQVTLNDIQLQM 272
DB  :|||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
DB  869 KRQYDQBIENLEKQKQTIERLBOEHTNRLRDAKRIKGEQEKEL--SKQNVLKN-----922
QY  273 EQHNRNSKLQRQENMELAEELKLIQ-----YELREE 305
DB  :|||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
DB  923 -RKKEQEFVQKQOQELDGLSKKIIIOQKAEALANIERECINNKKQMLNRAEAAIWELEER 981
QY  306 HIDKVFHKDLOQLOVDKILQQAQEMLKAE-----ERHQEKDPLLKEAV--ESQRMCELM 360
DB  :|||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
DB  982 HLQE--KHQLLKQALXDQYFMQHQHLLKHEKETEOMQRYNQVQNELIELKNRQTOEARLP 1039
QY  361 KQETHLKQOLALY-----TEKEPEP---QNTLSKSEVFTTFKQE-----398
DB  :|||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:

Mol. Cell. Biol. 20:2949-2949 (2000).
RN  [2]
RP  SEQUENCE FROM N.A.
RC  STRAIN=BALB/c;
RX  MEDLINE=20073281; PubMed=10611247;
RA  Sabourin L.A., Seale P., Wagner J., Rudnicki M.A.;
RT  "Caspase 3 cleavage of the Ste20-related kinase SLK releases and
XT  activates an apoptosis-inducing kinase domain and an actin-
XT  disassembling region.";
RL  Mol. Cell. Biol. 20:684-696(2000).
CC  -!- SIMILARITY: BELONGS TO THE SER/THR FAMILY OF PROTEIN KINASES.
DR  EMBL; AF112855; AAD28717.1; -.
DR  HSSP; P24941; LCKP.
DR  MGD; MGI:103241; Stk2.
DR  GO; GO:0005524; F:ATP binding; IEA.
DR  GO; GO:0003677; F:DNA binding; IEA.
DR  GO; GO:0004518; F:nuclease activity; IEA.
DR  GO; GO:0004674; F:protein serine/threonine kinase activity; IEA.
DR  GO; GO:0016740; F:transferase activity; IEA.
DR  GO; GO:0006289; P:nucleotide-excision repair; IEA.
DR  GO; GO:0006458; P:protein amino acid phosphorylation; IEA.
DR  InterPro; IPR000719; Prot_kinase.
DR  InterPro; IPR002290; Ser_thr_kinase.
DR  InterPro; IPR008271; Ser_thr_pkinase.
DR  InterPro; IPR001943; UvrB/C.
DR  Pfam; PF00069; pkinase; 1.
DR  ProDom; PD000001; Prot_kinase; 1.
DR  SMART; SM00220; S_TKC_1.
DR  PROSITE; PS00107; PROTEIN_KINASE_ATP; 1.
DR  PROSITE; PS00111; PROTEIN_KINASE_DOM; 1.
DR  PROSITE; PS00108; PROTEIN_KINASE_ST; 1.
DR  PROSITE; PS00151; UVR; 1.
KW  ATP-binding; Kinase; Serine/threonine-protein kinase; Transferase.
SQ  SEQUENCE 1202 AA; 137713 MW; 9F4C78A4C31FAAB2 CRC64;

Query Match          9.0%; Score 244; DB 11; Length 1212;
Best Local Similarity 25.1%; Pred. No. 0.00069;
Matches 114; Conservative 79; Mismatches 181; Indels 80; Gaps 17;

QY  75 NCGGFGDGAQGEPAEDAEKSRITVARNGEPEPTPVVYGEKPSKGPDPNTEIRQSDE 134
DB  :|||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
DB  5 NQG--GESSNGHVCPSK-----SIISDGGKGPSE----DAKKNKANRKEEDVNASGT 53
QY  135 VGDRDHRP-----QKKKAKAGLGE--ITLLMQTLTLSTPEEKLAALC--KTYAEILLE 185
DB  :|||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
DB  54 I--KHLPKPSGESEKTKSVLSKEDLIQLLSIMEGELQAREDVHMLRTEKTPEVLE 111
QY  186 EHRNSQKMKLIQ--KKOSQLVQEKDHLRGEHSKAVLARSKLESCLRELQOR---HNRSIK 240
DB  :|||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
DB  112 AHYSAEPEKVLRLVLRDAILAQEKSIGEDVYVEKPISELDRLERKQKETYRMLQLLIA 171
QY  241 BGVQVQARAREEERKKEVTSHPQVTLNDIQLQMEQHNRNSKLROE-----NMELAE 292
DB  :|||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
DB  172 EKCHRTVYELNEKHKHTDYAKSDDTNLLQERERLKKLLEQEKAYQARKEKENAKR 231
QY  293 LKKLLEQ-----YELREBHIDKVFHKDLOQLOVD---AKLQQAQEMLKAEABER 338
DB  :|||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
DB  232 LNKLDLDELVLKLSFALMLVDERQWHIEQL-----GLQSQKQVDLTQKLEEEBEKLVAVTYK 287
QY  339 HQREKDFLLKEAV-----ESQRMCELMQOETHLKO-----QLALYTEKPEEF 381
DB  :|||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
DB  288 SKEDQKLLKLEVDTEHTASRFSQHEEMNAKLQNSHNQRLKLVGLSORIELSET 347
QY  382 QNTLSKSEVFTTFKQEMKMTKKIKKLEKETTMYRSRWESSNKALLEMA---EKKTVRD 438
DB  :|||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
DB  348 NKSQKABE-----ELQELREKIAKCEGCGSSILMAEVESLRKRVLEMGKDEBITNTE 400
QY  439 KELEGLQVKIQRLKLCALQOTERNDLNKRVDL 472
DB  :|||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
DB  401 AQRELKKKQLEBEHHSKELRLEVEKQKRWSEL 434
```

RESULT 47
Q80U65 PRELIMINARY; PRT; 1307 AA.
AC Q80U65;
DT 01-JUN-2003 (TREMBlrel. 24, Created)
DT 01-JUN-2003 (TREMBlrel. 24, Last sequence update)
DT 01-OCT-2003 (TREMBlrel. 25, Last annotation update)
DE MKIAA0204 protein (Fragment).
GN MKIAA0204.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Brain;
RX MEDLINE=22579291; PubMed=12693553;
RA Okazaki N., Kikuno R., Ohara R., Inamoto S., Aizawa H., Yuasa S.,
RA Nakajima D., Nagase T., Ohara O., Koga H.;
RT "Prediction of the coding sequences of mouse homologues of KIAA genes:
RT II. The complete nucleotide sequences of 400 mouse KIAA-homologous
RT cDNAs identified by screening of terminal sequences of cDNA clones
RT randomly sampled from size-fractionated libraries.";
RL DNA Res. 10:35-48(2003).
DR EMBL; AK122218; BAC55500.1; -.
DR GO; GO:0005524; F:ATP binding; IEA.
DR GO; GO:0003677; F:DNA binding; IEA.
DR GO; GO:0004518; F:nuclease activity; IEA.
DR GO; GO:0004674; F:protein serine/threonine kinase activity; IEA.
DR GO; GO:0004713; F:protein-tyrosine kinase activity; IEA.
DR GO; GO:0006289; F:nucleotide-excision repair; IEA.
DR GO; GO:0006468; F:protein amino acid phosphorylation; IEA.
DR InterPro; IPR000719; Prot kinase.
DR InterPro; IPR002290; Ser Thr kinase.
DR InterPro; IPR008271; Ser Thr_pkin_AS.
DR InterPro; IPR001245; Tyr_pkinase.
DR InterPro; IPR001943; UvrB/C.
DR Pfam; PF00069; pkinase; 1.
DR ProDom; PD000001; Prot_kinase; 1.
DR SMART; SM00220; S_TKc; 1.
DR SMART; SM00219; TyKc; 1.
DR PROSITE; PS00107; PROTEIN KINASE ATP; 1.
DR PROSITE; PS00011; PROTEIN KINASE_DOM; 1.
DR PROSITE; PS00108; PROTEIN KINASE_ST; 1.
DR PROSITE; PS00151; UVR; 1.
FT NON_TER 1 1
SQ SEQUENCE 1307 AA; 148738 MW; 61178A1329DDCA24 CRC64;

Query Match 9.0%; Score 244; DB 11; Length 1307;
Best Local Similarity 21.2%; Pred. No. 0.00075;
Matches 133; Conservative 99; Mismatches 183; Indels 212; Gaps 28;

QY 6 QPEAGPE--GAQRRP-----SQAAPAV-EAEG--PGSQAPRKPEGAQARTASG 50
DB 687 KTEGPEAGAEHPGGRVEDKQPEQPAVCAVEAGQLTSTETTR---ATLQPEPTD 742

QY 51 ALRDVSEELS-RQLEDILTYVDNNGGPGEDG-----AQ 85
DB 743 EYEQVSESNISIELELVVITGAEARALGSEGAATAEVDLERKENAQVVPKASQAPAA 802

QY 86 GPAPAEPE-----DAKSTYVARNGEPTPVYGEKPSKG---DPNT----- 126
DB 803 SQPSEHPVLIPSIINSETTENKEEMGALPKPEILPPPEHEKGNUTDSTGTSVENS 862

QY 127 -----EETRQSEVGRDRHRRPQ--KKAKGL--GKEITLIMQTLNLTSTPE 170
DB 863 SGDLNLSIFLSFKANDSGSVLSQETTRQKTKLTKRKFIVDGEVSV---TTSKIVTDS 919

QY 171 EKLAALCKVYALLLEHRSNOKMKLQKKQ-----SOLVQEKDHLRGEHSKAVLAR 222
DB 920 D-----SKTEELRFLRRELRLQKEORAOQQLNGKLOQOOREQIFRRFEQEMLSK 973

QY 223 SK-----LESCLRELQRHNRSLKRGVORABEE-----EKKKEVTSHFOVTLNDIQLQM 272
DB 974 KFOYDOETENLEKQKQKQOTIERLEQSHTNRLRDEAKRIKGEQEKEL-SKFPQNVLKN----- 1027

QY 273 EQHNRNKLQRHNNMELAEKLLKLEQ-----YELREE 305
DB 1028 -RKKEEQFVQKQQLDGLSLKIIQQKAELANIERECLNNKQQLMRAREAIWELEER 1086

QY 306 HIDKVFKKDLQOQLVDLAKLQQAQEMLKAE--ERHOREKDFLKEAV--ESQRMCELM 360
DB 1087 HLQE--KHLLKQQLKQVFMQRHQLKRHEKETQMQRYNQRLIEELKNRQTOERARLP 1144

QY 361 KQETHLQKQALV-----TEKPEFP--ONTLSKSSEVFTTFKOE--- 398
DB 1145 KIORSEAKTRMAMFKSLRINSTATPDQDREKIKQFAAQBKKQKQRMQAQKHESQNR 1204

QY 399 -----MEQMTKKIKLEKETTMYRSWSSNKALLEMAEBK 434
DB 1205 DLQQLCEANVRELHQLQNEKCHLLVEHETQKLKELDEHSHOELKEWR-----EKL 1254

QY 435 TVRDELEGLQVKIQRLKLCRAQTE 461
DB 1255 RPKKTL-----BEPAKQLQEQ 1272

RESULT 48
Q922J3 PRELIMINARY; PRT; 1391 AA.
AC Q922J3;
DT 01-DEC-2001 (TREMBlrel. 19, Created)
DT 01-DEC-2001 (TREMBlrel. 19, Last sequence update)
DT 01-OCT-2003 (TREMBlrel. 25, Last annotation update)
DE Similar to restin (Reed-Steinberg cell-expressed intermediate
DE filament-associated protein).
GN RSN
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RA Strausberg R.;
RL Submitted (APR-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; BC007191; AAH07191.1; -.
DR MGD; MGI:1928401; RSN.
DR GO; GO:0005881; C:cytoplasmic microtubule; IDA.
DR GO; GO:0008017; F:microtubule binding; IDA.
DR GO; GO:0005515; F:protein binding; IPI.
DR InterPro; IPR000938; CAP-Gly.
DR InterPro; IPR001878; Znf_CCHC.
DR Pfam; PF01302; CAP_Gly; 2.
DR SMART; SM00343; Znf_C2HC; 1.
DR PROSITE; PS00845; CAP_GLY_1; 2.
DR PROSITE; PS02045; CAP_GLY_2; 2.
SQ SEQUENCE 1391 AA; 155813 MW; 061BED1FB3D4068D CRC64;

Query Match 9.0%; Score 244; DB 11; Length 1391;
Best Local Similarity 23.8%; Pred. No. 0.0008;
Matches 112; Conservative 86; Mismatches 153; Indels 120; Gaps 18;

QY 117 KEPSKGPNTETIRQSDVEGDR-----DHRPQKKKAKGLGKE-----ITLIMQTLN 164
DB 734 KANSEKLELTLLAQLEGAQKIKNLETENAESSKANSITKELQEKELVLTGLQDSL 793

QY 165 TLS-----TPBEKLAALCKVYALLLEHRSNOKM-----KLLQKKQ-----SOLVQEKD 209
DB 794 QNVQKTELEKELQTLKFKFASTSEBAVSAQTRMQDTVTKLHKEEQNVLSSELEKLE 853

QY 210 HLRGEHSK-----AVLRSKLSLCLRELQ--HNRS-----LKEEGVQR 246
DB 854 NLTDMAEKFKEDREBQLVKAEKLENDIAEINKMSGDSSQLTKWDELRLKERSVEE 913

QY 247 AREEEKKKEVTSHFOVTLNDIQLQMEQ-----HNERNKLQRHNNMELAEKLLKLE 298

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: June 7, 2004, 14:35:06 ; Search time 31 Seconds
(without alignments)
882.638 Million cell updates/sec

Title: US-10-023-529-8
Perfect score: 2702
Sequence: 1 KSSPQPGAGPEGAQRPSPQ.....APSTEASQGTGPQEPSTARA 530

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 389414 seqs, 51625971 residues

Total number of hits satisfying chosen parameters: 389414

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-Processing: Minimum Match 0%
Maximum Match 100%
Listing first 150 summaries

Database : Issued Patents AA:*
1: /cgn2_6/ptodata/2/iaa/5A_COMB.pep:*
2: /cgn2_6/ptodata/2/iaa/5B_COMB.pep:*
3: /cgn2_6/ptodata/2/iaa/5A_COMB.pep:*
4: /cgn2_6/ptodata/2/iaa/5B_COMB.pep:*
5: /cgn2_6/ptodata/2/iaa/PTUTUS_COMB.pep:*
6: /cgn2_6/ptodata/2/iaa/backfiles1.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2702	100.0	530	4	US-08-979-608A-8
2	2702	100.0	530	4	US-09-517-849-8
3	2702	100.0	530	4	US-09-616-289-8
4	2693	99.7	546	4	US-09-616-289-44
5	2465.5	91.2	557	4	US-08-979-608A-5
6	2465.5	91.2	557	4	US-09-517-849-5
7	2465.5	91.2	557	4	US-09-616-289-5
8	1038.5	38.4	386	2	US-08-968-751-6
9	248.5	9.2	786	4	US-09-688-188B-23
10	248.5	9.2	786	4	US-09-291-417D-23
11	248.5	9.2	1001	4	US-09-688-188B-31
12	248.5	9.2	1001	4	US-09-291-417D-31
13	247	9.1	787	4	US-09-688-188B-151
14	247	9.1	787	4	US-09-291-417D-151
15	247	9.1	1001	3	US-09-060-410-2
16	247	9.1	1001	4	US-09-723-458-2
17	240	8.9	2662	4	US-09-595-684B-31
18	239.5	8.9	1898	1	US-08-056-200-94
19	239.5	8.9	1898	4	US-08-800-644-94
20	236.5	8.8	1839	4	US-09-310-187A-1
21	231.5	8.6	1886	4	US-08-938-105-3
22	230.5	8.5	3878	4	US-09-914-259-11
23	228	8.4	1388	2	US-08-685-576-1
24	227.5	8.4	1162	2	US-08-728-323A-2
25	227.5	8.4	1162	4	US-09-298-568-2
26	227.5	8.4	1162	4	US-09-410-399-2
27	226.5	8.4	3248	1	US-08-353-700-1
28	226.5	8.4	3248	1	US-08-353-700-1
29	223	8.3	2101	1	US-08-195-487-4
30	223	8.3	2101	5	PCT-US93-06160-4
31	222	8.2	2101	1	US-08-466-390-4
32	222	8.2	2101	1	US-08-470-950-4
33	222	8.2	2101	1	US-08-467-781-4
34	222	8.2	2101	2	US-08-483-924-4
35	222	8.2	2101	3	US-09-452-294-1
36	220.5	8.2	2482	1	US-08-328-254-6
37	219.5	8.1	1695	4	US-09-866-108A-15753
38	218	8.1	443	2	US-08-795-475-6
39	217.5	8.0	967	4	US-09-914-259-21
40	217.5	8.0	1972	4	US-08-875-435B-4
41	215.5	8.0	816	2	US-08-533-306A-6
42	215.5	8.0	816	2	US-08-742-923A-6
43	215.5	8.0	885	2	US-08-533-306A-4
44	215.5	8.0	885	2	US-08-742-923A-4
45	215.5	8.0	1354	3	US-08-685-871-2
46	215	8.0	1388	4	US-09-976-594-296
47	214	7.9	1388	2	US-08-685-576-4
48	213	7.9	864	4	US-08-714-741-40
49	212	7.8	2954	4	US-09-150-867-1
50	211.5	7.8	1027	4	US-09-914-259-27
51	211	7.8	1581	4	US-09-866-108A-15754
52	207	7.7	977	4	US-09-010-147B-18
53	206.5	7.6	900	2	US-08-630-822A-62
54	206.5	7.6	900	2	US-09-005-069-62
55	206.5	7.6	900	4	US-09-171-156A-21
56	206.5	7.6	900	4	US-09-004-730A-21
57	206.5	7.6	900	4	US-08-981-799A-21
58	206.5	7.6	1530	4	US-09-976-594-736
59	206.5	7.6	1972	4	US-08-875-435B-3
60	206	7.6	472	4	US-09-166-350-17
61	206	7.6	611	4	US-09-216-393B-81
62	206	7.6	1044	4	US-09-107-532A-2229
63	206	7.6	1180	4	US-09-543-681A-6436
64	203.5	7.5	976	3	US-09-104-324B-4
65	203	7.5	2125	4	US-09-919-172-39
66	202	7.5	8991	4	US-08-714-741-32
67	201	7.4	748	4	US-08-688-188B-24
68	201	7.4	748	4	US-09-291-417D-24
69	200.5	7.4	534	4	US-09-103-664A-2
70	200	7.4	1231	4	US-08-714-741-41
71	199	7.4	458	4	US-08-529-055-73
72	198.5	7.3	576	2	US-08-533-306A-2
73	198.5	7.3	576	2	US-08-742-923A-2
74	198.5	7.3	1032	4	US-09-914-259-26
75	198	7.3	1279	4	US-09-724-517-2
76	198	7.3	1279	4	US-09-641-807A-2
77	198	7.3	1279	4	US-09-723-096-2
78	197.5	7.3	1618	1	US-07-853-913-4
79	197	7.3	993	3	US-09-060-410-4
80	197	7.3	993	4	US-09-723-458-4
81	196	7.3	966	4	US-09-688-188B-154
82	196	7.3	966	4	US-09-291-417D-154
83	195.5	7.2	898	4	US-09-688-188B-22
84	195.5	7.2	898	4	US-09-291-417D-22
85	195.5	7.2	1093	5	PCT-US93-03077-1
86	193.5	7.2	627	4	US-09-702-953B-3
87	193.5	7.2	647	4	US-09-702-953B-2
88	193.5	7.1	716	4	US-09-219-983A-20
89	192.5	7.1	1201	4	US-09-252-991A-32259
90	192	7.1	957	4	US-09-914-259-16
91	192	7.1	1388	4	US-09-572-191-2
92	192	7.1	1388	4	US-09-723-262-2
93	192	7.1	1388	4	US-09-723-219-2
94	191.5	7.1	456	4	US-09-808-701A-24
95	191.5	7.1	803	4	US-09-154-750A-85
96	191.5	7.1	803	4	US-09-665-479A-12
97	191.5	7.1	1298	4	US-09-645-456A-14
98	191.5	7.1	1298	4	US-09-425-324A-14
99	191.5	7.1	1298	4	US-09-645-791-14
100	191.5	7.1	1306	4	US-09-645-456A-10

101 191.5 7.1 1306 4 US-09-425-324A-10 Sequence 10, Appl
102 191.5 7.1 1306 4 US-09-645-791-10 Sequence 10, Appl
103 191 7.1 961 4 US-09-914-259-66 Sequence 66, Appl
104 191 7.1 968 4 US-09-688-188B-107 Sequence 107, Appl
105 191 7.1 968 4 US-09-688-188B-155 Sequence 155, Appl
106 191 7.1 968 4 US-09-291-417D-107 Sequence 107, Appl
107 191 7.1 968 4 US-09-291-417D-107 Sequence 107, Appl
108 191 7.1 2310 4 US-09-874-923-120 Sequence 120, Appl
109 191 7.1 2568 4 US-09-866-108A-3 Sequence 3, Appl
110 190.5 7.1 973 4 US-09-392-714-24 Sequence 24, Appl
111 190 7.0 595 4 US-09-462-951B-1 Sequence 1, Appl
112 190 7.0 779 4 US-10-164-595-56 Sequence 56, Appl
113 190 7.0 911 4 US-09-688-188B-92 Sequence 92, Appl
114 190 7.0 911 4 US-09-291-417D-92 Sequence 92, Appl
115 189.5 7.0 963 4 US-09-914-259-20 Sequence 20, Appl
116 189.5 7.0 1164 4 US-09-457-708-2 Sequence 2, Appl
117 189.5 7.0 1164 4 US-09-950-046A-2 Sequence 2, Appl
118 189.5 7.0 1164 4 US-09-976-594-989 Sequence 989, Appl
119 188 7.0 956 4 US-09-914-259-17 Sequence 17, Appl
120 187.5 6.9 953 4 US-09-914-259-22 Sequence 22, Appl
121 187 6.9 574 4 US-09-702-953B-4 Sequence 4, Appl
122 186 6.9 467 2 US-08-686-599A-17 Sequence 17, Appl
123 186 6.9 912 4 US-09-688-188B-26 Sequence 26, Appl
124 186 6.9 912 4 US-09-291-417D-26 Sequence 26, Appl
125 185.5 6.9 982 4 US-09-688-188B-90 Sequence 90, Appl
126 185.5 6.9 982 4 US-09-291-417D-90 Sequence 90, Appl
127 185 6.8 432 2 US-08-686-599A-18 Sequence 18, Appl
128 185 6.8 433 2 US-08-686-599A-5 Sequence 5, Appl
129 185 6.8 493 2 US-08-686-599A-16 Sequence 16, Appl
130 185 6.8 588 4 US-08-714-741-42 Sequence 42, Appl
131 185 6.8 683 6 5210183-3 Patent No. 5210183
132 184.5 6.8 959 4 US-09-914-259-67 Sequence 67, Appl
133 184.5 6.8 1031 4 US-09-914-259-24 Sequence 24, Appl
134 184.5 6.8 1312 2 US-08-686-599A-17 Sequence 17, Appl
135 184.5 6.8 1312 2 US-08-687-080-51 Sequence 51, Appl
136 184.5 6.8 1312 4 US-09-168-595-148 Sequence 148, Appl
137 184 6.8 591 4 US-09-252-991A-24904 Sequence 24904, A
138 183.5 6.8 704 4 US-09-370-838-191 Sequence 191, Appl
139 183.5 6.8 735 4 US-10-164-595-80 Sequence 80, Appl
140 183.5 6.8 784 4 US-10-164-595-79 Sequence 79, Appl
141 183.5 6.8 843 4 US-10-164-595-54 Sequence 54, Appl
142 183 6.8 1196 4 US-09-107-532A-3944 Sequence 3944, Ap
143 182 6.7 975 4 US-09-914-259-19 Sequence 19, Appl
144 182 6.7 1375 4 US-09-722-139-2 Sequence 2, Appl
145 182 6.7 1375 4 US-09-721-832-2 Sequence 2, Appl
146 182 6.7 1375 4 US-09-721-689-2 Sequence 2, Appl
147 181.5 6.7 1326 4 US-09-688-188B-15 Sequence 15, Appl
148 181.5 6.7 1326 4 US-09-291-417D-15 Sequence 15, Appl
149 181.5 6.7 1805 1 US-07-853-913-2 Sequence 2, Appl
150 181 6.7 468 4 US-09-328-352-6321 Sequence 6321, Ap

ALIGNMENTS

RESULT 1
US-08-979-608A-8
; Sequence 8, Application US/08979608A
; Patent No. 6355451
; GENERAL INFORMATION:
; APPLICANT: Lees, Ann M.
; Lees, Robert S.
; Iaw, Simon W.
; Arjona, Anibal A.
; TITLE OF INVENTION: NOVEL LOW DENSITY LIPOPROTEIN
; BINDING PROTEINS AND THEIR USES IN DIAGNOSING AND
; TREATING ATHEROSCLEROSIS
; NUMBER OF SEQUENCES: 42
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Fish & Richardson P.C.
; STREET: 225 Franklin Street
; CITY: Boston
; STATE: MA

COUNTRY: USA
ZIP: 02110-2804
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: FASTSEQ for Windows Version 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US 60/048,547
FILING DATE: 03-JUN-1997
APPLICATION NUMBER: US 60/031,930
FILING DATE: 27-NOV-1996
ATTORNEY/AGENT INFORMATION:
NAME: Myers, Louis
REGISTRATION NUMBER: 35,965
REFERENCE/DOCKET NUMBER: 10797-002001 (formerly 3983/59818)
TELECOMMUNICATION INFORMATION:
TELEPHONE: 617/542-5070
TELEFAX: 617/542-8906
INFORMATION FOR SEQ ID NO: 8:
SEQUENCE CHARACTERISTICS:
LENGTH: 530 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
SEQUENCE DESCRIPTION: SEQ ID NO: 8:
US-08-979-608A-8

Query Match 100.0%; Score 2702; DB 4; Length 530;
Best Local Similarity 100.0%; Pred. No. 2.2e-175;
Matches 530; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 KSSPOPEAGGACGAPSOAPAEAGPSSQAPRPGAGQAATAQSGALRDVSEELS 60
DB 1 KSSPOPEAGGACGAPSOAPAEAGPSSQAPRPGAGQAATAQSGALRDVSEELS 60
QY 61 RQLEDILSTYCVDDNQGGGEGDGAQGEPAEDAEKSRITYVARNGEPPTPVVYGEKPS 120
DB 61 RQLEDILSTYCVDDNQGGGEGDGAQGEPAEDAEKSRITYVARNGEPPTPVVYGEKPS 120
QY 121 KGPNTETIROSDVGDHRRPQKKAKGLGKEITLLMQTLNTLSTPEEKLAALCKKY 180
DB 121 KGPNTETIROSDVGDHRRPQKKAKGLGKEITLLMQTLNTLSTPEEKLAALCKKY 180
QY 181 AELLBEHRNSQOMKLLQKQSQQLVQKDHLEHSHKAVLARSKLESCLREIQRHNRSLK 240
DB 181 AELLBEHRNSQOMKLLQKQSQQLVQKDHLEHSHKAVLARSKLESCLREIQRHNRSLK 240
QY 241 REGVQAREEERKREKVTSHFOVTLNDIQLQMEQHNERNNSKLQRQNMELAEKLKLEQY 300
DB 241 REGVQAREEERKREKVTSHFOVTLNDIQLQMEQHNERNNSKLQRQNMELAEKLKLEQY 300
QY 301 EUREEHIDKVFHKDLQQLVDKQLQQAQEMLKEAEERHOREKDFLLKEAVESQRCMLM 360
DB 301 EUREEHIDKVFHKDLQQLVDKQLQQAQEMLKEAEERHOREKDFLLKEAVESQRCMLM 360
QY 361 KQOETHLQKQALALYTERFEFQNTLSKSEVFTTFKQEMEKMTKKIKKLEKETTYRSRW 420
DB 361 KQOETHLQKQALALYTERFEFQNTLSKSEVFTTFKQEMEKMTKKIKKLEKETTYRSRW 420
QY 421 ESSNKALLEMAEKTVRDKELEGLQVKIQRLEKCRALQTERNDLNKRVDLSAGQGSLL 480
DB 421 ESSNKALLEMAEKTVRDKELEGLQVKIQRLEKCRALQTERNDLNKRVDLSAGQGSLL 480
QY 481 TDSGPERPEGGAQAPSPRVTEAPCYPGAPSTASGQTGPQEPPTSARA 530
DB 481 TDSGPERPEGGAQAPSPRVTEAPCYPGAPSTASGQTGPQEPPTSARA 530

US-09-517-849-8

; Sequence 8, Application US/09517849
; Patent No. 6605588
; GENERAL INFORMATION:
; APPLICANT: Lees, Ann M.
; Lees, Robert S.
; Law, Simon W.
; Arjona, Anibal A.

TITLE OF INVENTION: NOVEL LOW DENSITY LIPOPROTEIN

; BINDING PROTEINS AND THEIR USES IN DIAGNOSING AND
; TREATING ATHEROSCLEROSIS

NUMBER OF SEQUENCES: 42

; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Fish & Richardson P.C.
; STREET: 225 Franklin Street
; CITY: Boston
; STATE: MA
; COUNTRY: USA
; ZIP: 02110-2804

COMPUTER READABLE FORM:

; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS

SOFTWARE: FastSeq for Windows Version 2.0

; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/517,849
; FILING DATE: 02-Mar-2000

PRIOR APPLICATION DATA:

; APPLICATION NUMBER: 08/979,608
; FILING DATE: 26-NOV-1997

ATTORNEY/AGENT INFORMATION:

; NAME: Myers, Louis
; REGISTRATION NUMBER: 35,965
; REFERENCE/DOCKET NUMBER: 10797-003001

TELECOMMUNICATION INFORMATION:

; TELEPHONE: 617/542-5070
; TELEFAX: 617/542-8906

INFORMATION FOR SEQ ID NO: 8:

; SEQUENCE CHARACTERISTICS:
; LENGTH: 530 amino acids
; TYPE: amino acid
; TOPOLOGY: linear

MOLECULE TYPE: protein

; SEQUENCE DESCRIPTION: SEQ ID NO: 8:

US-09-517-849-8

Query Match 100.0%; Score 2702; DB 4; Length 530;

Best Local Similarity 100.0%; Pred. No. 2.2e-175;

Matches 530; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 KSSPGQPEAGPGAGQERPSQAAPAVEAEGPGSSQAAPRKEGQAQARTAQSGALRDVSELS 60

Db 1 KSSPGQPEAGPGAGQERPSQAAPAVEAEGPGSSQAAPRKEGQAQARTAQSGALRDVSELS 60

Qy 61 RQLEDILSTYCVNNQGGPGEDGAQGEPAEPDAEKSRITYVARNGEPEPTPVVYGEKPS 120

Db 61 RQLEDILSTYCVNNQGGPGEDGAQGEPAEPDAEKSRITYVARNGEPEPTPVVYGEKPS 120

Qy 121 KGDPTNTEIRQSDVGDGRDHRPQEKKAGLGKEITLLMOTLNTLSTPEKLAALCKKY 180

Db 121 KGDPTNTEIRQSDVGDGRDHRPQEKKAGLGKEITLLMOTLNTLSTPEKLAALCKKY 180

Qy 181 AELLEHRNSQKMKLLQKKQSQLVQEKDHLRGHSHKAVLARSKLESICRELQHRNLSK 240

Db 181 AELLEHRNSQKMKLLQKKQSQLVQEKDHLRGHSHKAVLARSKLESICRELQHRNLSK 240

Qy 241 EGVQVRAEEERKEKVTSHFQVTLNDIQLOMEQHNRNSKLRQENMELARLKKLIEQY 300

Db 241 EGVQVRAEEERKEKVTSHFQVTLNDIQLOMEQHNRNSKLRQENMELARLKKLIEQY 300

Qy 301 ELREEHIDKVFKHDLQQLQQLVDAKLQQAQEMLKEAERHOREKDFLLKEAVESQRMCELM 360

Db 301 ELREEHIDKVFKHDLQQLQQLVDAKLQQAQEMLKEAERHOREKDFLLKEAVESQRMCELM 360

Qy 361 KOETHLKOQLALYTEKEFEFQNTLSKSEVPTTFKQEMEKMTKKIKLEKETTMYRGRW 420

Db 361 KOETHLKOQLALYTEKEFEFQNTLSKSEVPTTFKQEMEKMTKKIKLEKETTMYRGRW 420

Qy 421 ESSNKALLEMAEETKTVRDKEGLEQLVKIQRIQLEKLCRALQTERNDLNKRVQDLSAGQGS 480

Db 421 ESSNKALLEMAEETKTVRDKEGLEQLVKIQRIQLEKLCRALQTERNDLNKRVQDLSAGQGS 480

Qy 481 TDSGPERPEPGAGQAPSSPRVTEAPCYPGAPSTASQOTGPQEBTSARA 530

Db 481 TDSGPERPEPGAGQAPSSPRVTEAPCYPGAPSTASQOTGPQEBTSARA 530

RESULT 3

US-09-616-289-8

; Sequence 8, Application US/09616289

; Patent No. 6632923

; GENERAL INFORMATION:

; APPLICANT: Lees, Ann M.

; APPLICANT: Lees, Robert S.

; APPLICANT: Law, Simon W.

; APPLICANT: Arjona, Anibal A.

; TITLE OF INVENTION: NOVEL LOW DENSITY LIPOPROTEIN BINDING

; TITLE OF INVENTION: PROTEINS AND THEIR USE IN DIAGNOSING AND TREATING

; TITLE OF INVENTION: ATHEROSCLEROSIS

; FILE REFERENCE: 10797-004001

; CURRENT APPLICATION NUMBER: US/09/616,289

; CURRENT FILING DATE: 2000-07-14

; PRIOR APPLICATION NUMBER: US 09/517,849

; PRIOR FILING DATE: 2000-03-02

; PRIOR APPLICATION NUMBER: US 08/979,608

; PRIOR FILING DATE: 1997-11-26

; PRIOR APPLICATION NUMBER: US 60/031,930

; PRIOR FILING DATE: 1996-11-27

; PRIOR APPLICATION NUMBER: US 60/048,547

; PRIOR FILING DATE: 1997-06-03

; NUMBER OF SEQ ID NOS: 53

; SOFTWARE: FastSeq for Windows Version 4.0

; SEQ ID NO 8

; LENGTH: 530

; TYPE: PRT

; ORGANISM: Homo sapiens

US-09-616-289-8

Query Match 100.0%; Score 2702; DB 4; Length 530;

Best Local Similarity 100.0%; Pred. No. 2.2e-175;

Matches 530; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 KSSPGQPEAGPGAGQERPSQAAPAVEAEGPGSSQAAPRKEGQAQARTAQSGALRDVSELS 60

Db 1 KSSPGQPEAGPGAGQERPSQAAPAVEAEGPGSSQAAPRKEGQAQARTAQSGALRDVSELS 60

Qy 61 RQLEDILSTYCVNNQGGPGEDGAQGEPAEPDAEKSRITYVARNGEPEPTPVVYGEKPS 120

Db 61 RQLEDILSTYCVNNQGGPGEDGAQGEPAEPDAEKSRITYVARNGEPEPTPVVYGEKPS 120

Qy 121 KGDPTNTEIRQSDVGDGRDHRPQEKKAGLGKEITLLMOTLNTLSTPEKLAALCKKY 180

Db 121 KGDPTNTEIRQSDVGDGRDHRPQEKKAGLGKEITLLMOTLNTLSTPEKLAALCKKY 180

Qy 181 AELLEHRNSQKMKLLQKKQSQLVQEKDHLRGHSHKAVLARSKLESICRELQHRNLSK 240

Db 181 AELLEHRNSQKMKLLQKKQSQLVQEKDHLRGHSHKAVLARSKLESICRELQHRNLSK 240

Qy 241 EGVQVRAEEERKEKVTSHFQVTLNDIQLOMEQHNRNSKLRQENMELARLKKLIEQY 300

Db 241 EGVQVRAEEERKEKVTSHFQVTLNDIQLOMEQHNRNSKLRQENMELARLKKLIEQY 300

Qy 301 ELREEHIDKVFKHDLQQLQQLVDAKLQQAQEMLKEAERHOREKDFLLKEAVESQRMCELM 360

Db 301 ELREEHIDKVFKHDLQQLQQLVDAKLQQAQEMLKEAERHOREKDFLLKEAVESQRMCELM 360

QY 361 KQETHLKQOALALYTERFEFQNTLSKSSSEVFTTFKQEMERWTKKIKKLEKETTYRSRW 420
Db 361 KQETHLKQOALALYTERFEFQNTLSKSSSEVFTTFKQEMERWTKKIKKLEKETTYRSRW 420
QY 421 ESSNKALLEMAEKEKTVRDKEGLQVKIORLEKLCRALQTERNDLNKRVODLSAGQGSGL 480
Db 421 ESSNKALLEMAEKEKTVRDKEGLQVKIORLEKLCRALQTERNDLNKRVODLSAGQGSGL 480
QY 481 TDSGPERPREGGAQAPSSPRVTEAPCYPCAPSTASGQTGPQPTTSARA 530
Db 481 TDSGPERPREGGAQAPSSPRVTEAPCYPCAPSTASGQTGPQPTTSARA 530

RESULT 4

US-09-616-289-44
; Sequence 44, Application US/09616289
; Patent No. 6632923
; GENERAL INFORMATION:
; APPLICANT: Lees, Ann M.
; APPLICANT: Lees, Robert S.
; APPLICANT: Law, Simon W.
; APPLICANT: Arjona, Anibal A.
; TITLE OF INVENTION: NOVEL LOW DENSITY LIPOPROTEIN BINDING
; TITLE OF INVENTION: PROTEINS AND THEIR USE IN DIAGNOSING AND TREATING
; TITLE OF INVENTION: ATHEROSCLEROSIS
; FILE REFERENCE: 10797-004001
; CURRENT APPLICATION NUMBER: US/09/616,289
; CURRENT FILING DATE: 2000-07-14
; PRIOR FILING DATE: 2000-03-02
; PRIOR FILING DATE: 2000-03-02
; PRIOR FILING DATE: 1997-11-26
; PRIOR FILING DATE: 1996-11-27
; PRIOR FILING DATE: 1997-06-03
; NUMBER OF SEQ ID NOS: 53
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 44
; LENGTH: 546
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-616-289-44

Query Match 99.7%; Score 2693; DB 4; Length 546;
Best Local Similarity 99.8%; Pred. No. 9.3e-175;
Matches 529; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 KSSPGQPEAGPEGAQERPSQAAPAVEABGPGSSQAAPKPEGAQARTAQSGALRDVSEELS 60
Db 17 KSSPGQPEAGPEGAQERPSQAAPAVEABGPGSSQAAPKPEGAQARTAQSGALRDVSEELS 76
QY 61 RQLEDILSTYCVDNQGGPGEDGAQGPAPPEPAEKSRITVARNGEPEPTPVVYVGEKPS 120
Db 77 RQLEDILSTYCVDNQGGPGEDGAQGPAPPEPAEKSRITVARNGEPEPTPVVYVGEKPS 136
QY 121 KGDPTNTEIRQSDVEGDRHRRPOEKKAKGLGKEITLLMOTLNTLSTPEEKLAALCKKY 180
Db 137 KGDPTNTEIRQSDVEGDRHRRPOEKKAKGLGKEITLLMOTLNTLSTPEEKLAALCKKY 196
QY 181 AELLEERHSKQMKLLQKQSQOLVQKHLRGHSHKAVLARSKLSLCRELQRHNSLK 240
Db 197 AELLEERHSKQMKLLQKQSQOLVQKHLRGHSHKAVLARSKLSLCRELQRHNSLK 256
QY 241 EGVQARBEERKEKVTSHFQVTLNDIQLOMEQHNRHSKLRQENMELAEKLLIEQY 300
Db 257 EGVQARBEERKEKVTSHFQVTLNDIQLOMEQHNRHSKLRQENMELAEKLLIEQY 316
QY 301 ELREEHIDKVFHKDIQQQLVDKQLQQAQEMLEKEAERHOREKDFLLKEAVESQRCMEL 360
Db 317 ELREEHIDKVFHKDIQQQLVDKQLQQAQEMLEKEAERHOREKDFLLKEAVESQRCMEL 376
QY 361 KQETHLKQOALALYTERFEFQNTLSKSSSEVFTTFKQEMERWTKKIKKLEKETTYRSRW 420

Db 377 KQETHLKQOALALYTERFEFQNTLSKSSSEVFTTFKQEMERWTKKIKKLEKETTYRSRW 436
QY 421 ESSNKALLEMAEKEKTVRDKEGLQVKIORLEKLCRALQTERNDLNKRVODLSAGQGSGL 480
Db 437 ESSNKALLEMAEKEKTVRDKEGLQVKIORLEKLCRALQTERNDLNKRVODLSAGQGSGL 496
QY 481 TDSGPERPREGGAQAPSSPRVTEAPCYPCAPSTASGQTGPQPTTSARA 530
Db 497 TDSGPERPREGGAQAPSSPRVTEAPCYPCAPSTASGQTGPQPTTSARA 546

RESULT 5

US-08-979-608A-5
; Sequence 5, Application US/08979608A
; Patent No. 6355451
; GENERAL INFORMATION:
; APPLICANT: Lees, Ann M.
; APPLICANT: Lees, Robert S.
; APPLICANT: Law, Simon W.
; APPLICANT: Arjona, Anibal A.
; TITLE OF INVENTION: NOVEL LOW DENSITY LIPOPROTEIN
; TITLE OF INVENTION: BINDING PROTEINS AND THEIR USES IN DIAGNOSING AND
; TITLE OF INVENTION: TREATING ATHEROSCLEROSIS
; NUMBER OF SEQUENCES: 42
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Fish & Richardson P.C.
; STREET: 225 Franklin Street
; CITY: Boston
; STATE: MA
; COUNTRY: USA
; ZIP: 02110-2804
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq for Windows Version 2.0
; CURRENT APPLICATION DATA: US/08/979,608A
; FILING DATE: 26-NOV-1997
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 60/048,547
; FILING DATE: 03-JUN-1997
; APPLICATION NUMBER: US 60/031,930
; FILING DATE: 27-NOV-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: Myers, Louis
; REGISTRATION NUMBER: 35,965
; REFERENCE/DOCKET NUMBER: 10797-002001 (formerly 3983/59818)
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 617/542-5070
; TELEFAX: 617/542-8906
; INFORMATION FOR SEQ ID NO: 5:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 557 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; FRAGMENT TYPE: internal
; SEQUENCE DESCRIPTION: SEQ ID NO: 5:
US-08-979-608A-5

Query Match 91.2%; Score 2465.5; DB 4; Length 557;
Best Local Similarity 90.8%; Pred. No. 2.5e-159;
Matches 492; Conservative 12; Mismatches 25; Indels 13; Gaps 3;

QY 1 KSSPGQPEAGPEGAQERPSQAAPAVEABGPGSSQAAPKPEGAQARTAQSGALRDVSEELS 60
Db 17 KSSPGQPEAGPEGAQERPSQAAPAVEABGPGSSQAAPKPEGAQARTAQSGALRDVSEELS 75
QY 61 RQLEDILSTYCVDNQGGPGEDGAQGPAPPEPAEKSRITVARNGEPEPTPVVYVGEKPS 119
Db 76 RQLEDILSTYCVDNQGGPGEDGAQGPAPPEPAEKSRITVARNGEPEPTPVVYVGEKPS 135

Qy	120	SKGDPNTEET	ROSDEVDGR	HRBPBKKA	GLGKEIT	LLMOT	NTLST	TPPEKLAAL	CKK	179
		:	:	:	:	:	:	:	:	
Db	136	SKAAPTET	EIRTSDEVDGR	HRBPBKKA	GLGKEIT	LLMOT	NTLST	TPPEKLAAL	CKK	195
		:	:	:	:	:	:	:	:	
Qy	180	YAEFLAEHNS	QOMKLLQKQ	KSQVQ	VEQDKHLR	GEHSAVLARS	KLSESLCRL	QENRSL	239	
		:	:	:	:	:	:	:	:	
Db	196	YAEFLAEHNS	QOMKLLQKQ	KSQVQ	VEQDKHLR	GEHSAKILLARS	KLSESLCRL	QENRSL	255	
		:	:	:	:	:	:	:	:	
Qy	240	KEEGVORABE	BEKKVETSH	EQVTLNDI	OLQMHNRNS	KLROENMEL	AEPLK	KLIEQ	299	
		:	:	:	:	:	:	:	:	
Db	256	KEEGVORABE	BEKKVETSH	EQVTLNDI	OLQMHNRNS	KLROENMEL	AEPLK	KLIEQ	315	
		:	:	:	:	:	:	:	:	
Qy	300	YELREBHI	DKVPKHKD	LQQLVDA	KLOQAQEK	LKEABER	HOREK	DPLLEAVESQ	RMCFL	359
		:	:	:	:	:	:	:	:	
Db	316	YELREBHI	DKVPKHKD	LQQLVDA	KLOQAQEK	LKEABER	HOREK	DPLLEAVESQ	RMCFL	375
		:	:	:	:	:	:	:	:	
Qy	360	MKQOETHL	KQOLALY	TEKPEEP	QNTLSK	SSVPTTFQ	EMEKMTKKI	KLEKETTMY	R 419	
		:	:	:	:	:	:	:	:	
Db	376	MKQOETHL	KQOLALY	TEKPEEP	QNTLSK	SSVPTTFQ	EMEKMTKKI	KLEKETTMY	R 435	
		:	:	:	:	:	:	:	:	
Qy	420	WESSNKALL	EMAEKTV	RDKLEGL	QVXIO	RLEKLCRAL	QTERND	LNKRVODLS	SAGQGS	479
		:	:	:	:	:	:	:	:	
Db	436	WESSNKALL	EMAEKTV	RDKLEGL	QVXIO	RLEKLCRAL	QTERND	LNKRVODLS	SAGQGP	495
		:	:	:	:	:	:	:	:	
Qy	480	LTDGSPERP	-----	BGCAQAP	SSPRVTE	APCYFG	APSTFAS	QOTGQ	SPPTSA	528
		:	:	:	:	:	:	:	:	
Db	496	VSDGSPERP	SPATTSK	QGVGBG	PAQVNP	SRATDAS	CCAGAP	STFASQ	TGQBPPTTA	555
		:	:	:	:	:	:	:	:	
Qy	529	RA	530							
Db	556	TA	557							

RESULT 6

US-09-517-849-5
: Sequence 5, Application US/09517849
: Patent No. 6605588
: GENERAL INFORMATION:
: APPLICANT: Lees, Ann M.
: Lees, Robert S.
: Law, Simon W.
: Arjona, Anibal A.
: TITLE OF INVENTION: NOVEL LOW DENSITY LIPOPROTEIN
: BINDING PROTEINS AND THEIR USES IN DIAGNOSING AND
: TREATING ATHEROSCLEROSIS
: NUMBER OF SEQUENCES: 42
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Fish & Richardson P.C.
: STREET: 225 Franklin Street
: CITY: Boston
: STATE: MA
: COUNTRY: USA
: ZIP: 02110-2804
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Diskette
: COMPUTER: IBM Compatible
: OPERATING SYSTEM: DOS
: SOFTWARE: FastSeq for Windows Version 2.0
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/09/517,849
: FILING DATE: 02-Mar-2000
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: 08/979,608
: FILING DATE: 26-NOV-1997
: ATTORNEY/AGENT INFORMATION:
: NAME: Myers, Louis
: REGISTRATION NUMBER: 35,965
: REFERENCE/DOCKET NUMBER: 10797-003001
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: 617/542-5070
: TELEFAX: 617/542-8906
: INFORMATION FOR SEQ ID NO: 5:

```

1  SEQUENCE CHARACTERISTICS:
2  LENGTH: 557 amino acids
3  TYPE: amino acid
4  TOPOLOGY: linear
5  MOLECULE TYPE: protein
6  FRAGMENT TYPE: internal
7  SEQUENCE DESCRIPTION: SEQ ID NO: 5:
8  US-09-517-849-5

```

Query Match	91.2%;	Score	2465.5;	DB	4;	Length	557;
Best Local Similarity	90.8%;	Pred.	NO	2.5e-159;			
Matches	492;	Conservative	12;	Mismatches	25;	Indels	13;
							Gaps
							3;
Qy	1	KSSPGQPEAGPEGAQERPSQAAPAVAEAGPVGSSQAQPRKPEGAQAQTAQSGALRDVSEELS	60				
Db	17	KSSPGQPEAGPEGAQERPSQAAPAVAEAGPVGSSQAQPRKPEGAQAQTAQSGALRDVSEELS	75				
Qy	61	RLQEDILSYCYDNNOGGDEGAQGEPAEPEDAEKSRITYVAENGEPEP	119				
Db	76	RLQEDILSYCYDNNOGGDEGAQGEVQGEPEPEPDAEKSRAYVARGNEPEPGTVPVNGEKT	135				
Qy	120	SKGDPNTEBIROSDEYGDHRRPOEKKAKGLGKEITILLMQTLNTLSTPBEKLAALCKK	179				
Db	136	SKAEPGTBEIRTDEYGDHRRPOEKKAKGLGKEITILLMQTLNTLSTPBEKLAALCKK	195				
Qy	180	YAEELLEHNRSOKMLLOKQOSOLVOEKDHLRGHSKAVLARSKLESICRELQHNRSI	239				
Db	196	YAEELLEHNRSOKMLLOKQOSOLVOEKDHLRGHSKAILARSKLESICRELQHNRSI	255				
Qy	240	XBEGVQRAEERBEKKREVTSHFQVTLNDIQLQWQHNRNSKLRQENMELARLKKLIDQ	299				
Db	256	XBEGVQRAEERBEKKREVTSHFQVTLNDIQLQWQHNRNSKLRQENMELARLKKLIDQ	315				
Qy	300	YELREBHIKVFQKHDLQOOLVDAKLQQAQEMLKEABERHQREKDFLLKEAVESQWCEL	359				
Db	316	YELREBHIKVFQKHDLQOOLVDAKLQQAQEMLKEABERHQREKDFLLKEAVESQWCEL	375				
Qy	360	MKGQETHLQOOLALYTEKEFEFQNTLSKSSVPTFTTKQEMEKMTKKIKKLEKETTMYRSR	419				
Db	376	MKGQETHLQOOLALYTEKEFEFQNTLSKSSVPTFTTKQEMEKMTKKIKKLEKETTMYRSR	435				
Qy	420	WESSNKALLEMABEKTVRDKLEGLQVKTORLEKLCRALQTERNDLNKRVQDLSAGGQGS	479				
Db	436	WESSNKALLEMABEKTVRDKLEGLQVKTORLEKLCRALQTERNDLNKRVQDLSAGGQGP	495				
Qy	480	LTDSGERRP-----EGPGQAQPSRRVTEAPCYQAPSTEAAGQTGPQPTSA	528				
Db	496	VSDSGERRPEPATTSKEQGVGPGQAQVPNSPRAVDASCCAGAPSTEAAGQTGPQPTTA	555				
Qy	529	RA	530				
Db	556	TA	557				

RESULT 7

```

US-09-816-289-5
; Sequence 5, Application US/09616289
; Patent No. 6632923
; GENERAL INFORMATION:
; APPLICANT: Lees, Ann M.
; APPLICANT: Lees, Robert S.
; APPLICANT: Law, Simon W.
; APPLICANT: Arjona, Anibal A.
; TITLE OF INVENTION: NOVEL LOW DENSITY LIPOPROTEIN BINDING
; TITLE OF INVENTION: PROTEINS AND THEIR USE IN DIAGNOSING AND TREATING
; TITLE OF INVENTION: ATHEROSCLEROSIS
; FILE REFERENCE: 10797-004001
; CURRENT APPLICATION NUMBER: US/09/616,289
; CURRENT FILING DATE: 2000-07-14
; PRIOR APPLICATION NUMBER: US 09/517,849
; PRIOR FILING DATE: 2000-03-02
; PRIOR APPLICATION NUMBER: US 08/979,608
; PRIOR FILING DATE: 1997-11-26

```


; NUMBER OF SEQ ID NOS: 155
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 23
; LENGTH: 786
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-291-417D-23

Query Match
Best Local Similarity 21.6%; Score 248.5; DB 4; Length 786;
Matches 139; Conservative 117; Mismatches 197; Indels 191; Gaps 29;
QY 23 PAVEAE-----GPGSSQA-PRKPEGAQARTAGSALRDVSEHLSRQLE 64
DB 109 PAVEAQEEEEQDHGVRGTGTVNSVGSNQSIFSMSSISASSQSSVNSLFDVDDKS-ELD 167
QY 65 DILSTCYVDNNOG---GPGEDG--AQGEPAEPEDAERSTTVARNGEPEPTPVVYGEKE 118
DB 168 MMEGDHTVMSNVSIVHLKPEENYREGDP-----RT---RASDPQSPQVSRHKS 215
QY 119 PSKGDPTTEIRQSDVGDGRDRRPPQEKKAKGLGKEIT---LLMQTLNTLSTPEKLA 174
DB 216 HYRNREHFATIRTASLV---TRQMEHQDSSELREQMSGYKRMRRQHQKQMLTLENK 271
QY 175 ALCKKYA---ELLEHEHN--SOKMKLQKQSOVLQV-----KDLRGEHSK 217
DB 272 AEMDEHRLDKDLETQNNFAAEMKLIKQQAAMEKAKVMSNEEKKFQOHIOAQOKK 331
QY 218 AVLARSKLSLCELRQHRNLSKEGVQV---AREEEE---KRKEVTSHTFQV--TLNDI 268
DB 332 EL--NSFLESQKREYKLRKEQLKBELENQSTPKKEKQEWLSKQKENIQHFOABEENLL 389
QY 269 QLQME-----OHN-----ERNSKLRQENMELAERLKLIEQVELRE 304
DB 390 RRQRYLELECRPKRRMLLGRHNLQDLVRELNKRTQKDLFAHMLLRQHESMQLF 449
QY 305 EHIDKVPKHDLOQOLVDKLAQQAQEMLEAEERHOREKDFLLKEAVESQRMCELMKQOE 364
DB 450 RHLNTIQK---MRCELI--RLQHOTELTNQLEYNKRRERELRRKXVMEVRSQPKSLKSE 504
QY 365 THLKOQL---ALYTEKPEFQNTLSKSEVFTTFQKQEMKNTXIKKLEKETTYRSRW 420
DB 505 LQIKKQFQDTCKIQRYKALRNHLE---TTPKSEKAVLKLK--EQTRKLAULA 557
QY 421 ESSNKALLEMAEKTVR-----DKBLE 442
DB 558 EYDHSINEMLSQALRLDEAQAECQVLKMQLOLELLNAYQSKIKMQABAQHDREL 617
QY 443 GLQVKI-----QRLEKLCRALQTERND-----LNKRVODLSAGGQGS-----L 480
DB 618 ELEQVSLRRALLEQKIEEMLALQNERTERISLLEQAREIEAFDSMELGFSNMVL 677
QY 481 TDSGPRRPEGPAQAPSPRVTEAPCYGAP--STEASGQTGP 522
DB 678 SNLSPE-----AFSHSYPGASGWSHNPTGGPGP 705

RESULT 11
US-09-188B-31
; Sequence 31, Application US/09688188B
; Patent No. 6856716
; GENERAL INFORMATION:
; APPLICANT: PLOWMAN, GREGORY
; APPLICANT: MARTINEZ, RICARDO
; APPLICANT: WHYTE, DAVID
; TITLE OF INVENTION: STE20-RELATED PROTEIN KINASES
; FILE REFERENCE: 038602/0328
; CURRENT APPLICATION NUMBER: US/09/688,188B
; CURRENT FILING DATE: 2000-10-16
; PRIOR APPLICATION NUMBER: 09/291,417
; PRIOR FILING DATE: 1999-04-14
; PRIOR APPLICATION NUMBER: 60/081,784
; PRIOR FILING DATE: 1998-04-14

; PRIOR APPLICATION NUMBER: 60/081,784
; PRIOR FILING DATE: 1998-04-14
; NUMBER OF SEQ ID NOS: 155
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 23
; LENGTH: 786
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-688-188B-23

Query Match
Best Local Similarity 21.6%; Score 248.5; DB 4; Length 786;
Matches 139; Conservative 117; Mismatches 197; Indels 191; Gaps 29;
QY 23 PAVEAE-----GPGSSQA-PRKPEGAQARTAGSALRDVSEHLSRQLE 64
DB 109 PAVEAQEEEEQDHGVRGTGTVNSVGSNQSIFSMSSISASSQSSVNSLFDVDDKS-ELD 167
QY 65 DILSTCYVDNNOG---GPGEDG--AQGEPAEPEDAERSTTVARNGEPEPTPVVYGEKE 118
DB 168 MMEGDHTVMSNVSIVHLKPEENYREGDP-----RT---RASDPQSPQVSRHKS 215
QY 119 PSKGDPTTEIRQSDVGDGRDRRPPQEKKAKGLGKEIT---LLMQTLNTLSTPEKLA 174
DB 216 HYRNREHFATIRTASLV---TRQMEHQDSSELREQMSGYKRMRRQHQKQMLTLENK 271
QY 175 ALCKKYA---ELLEHEHN--SOKMKLQKQSOVLQV-----KDLRGEHSK 217
DB 272 AEMDEHRLDKDLETQNNFAAEMKLIKQQAAMEKAKVMSNEEKKFQOHIOAQOKK 331
QY 218 AVLARSKLSLCELRQHRNLSKEGVQV---AREEEE---KRKEVTSHTFQV--TLNDI 268
DB 332 EL--NSFLESQKREYKLRKEQLKBELENQSTPKKEKQEWLSKQKENIQHFOABEENLL 389
QY 269 QLQME-----OHN-----ERNSKLRQENMELAERLKLIEQVELRE 304
DB 390 RRQRYLELECRPKRRMLLGRHNLQDLVRELNKRTQKDLFAHMLLRQHESMQLF 449
QY 305 EHIDKVPKHDLOQOLVDKLAQQAQEMLEAEERHOREKDFLLKEAVESQRMCELMKQOE 364
DB 450 RHLNTIQK---MRCELI--RLQHOTELTNQLEYNKRRERELRRKXVMEVRSQPKSLKSE 504
QY 365 THLKOQL---ALYTEKPEFQNTLSKSEVFTTFQKQEMKNTXIKKLEKETTYRSRW 420
DB 505 LQIKKQFQDTCKIQRYKALRNHLE---TTPKSEKAVLKLK--EQTRKLAULA 557
QY 421 ESSNKALLEMAEKTVR-----DKBLE 442
DB 558 EYDHSINEMLSQALRLDEAQAECQVLKMQLOLELLNAYQSKIKMQABAQHDREL 617
QY 443 GLQVKI-----QRLEKLCRALQTERND-----LNKRVODLSAGGQGS-----L 480
DB 618 ELEQVSLRRALLEQKIEEMLALQNERTERISLLEQAREIEAFDSMELGFSNMVL 677
QY 481 TDSGPRRPEGPAQAPSPRVTEAPCYGAP--STEASGQTGP 522
DB 678 SNLSPE-----AFSHSYPGASGWSHNPTGGPGP 705

RESULT 10
US-09-291-417D-23
; Sequence 23, Application US/09291417D
; Patent No. 680170
; GENERAL INFORMATION:
; APPLICANT: PLOWMAN, GREGORY
; APPLICANT: MARTINEZ, RICARDO
; APPLICANT: WHYTE, DAVID
; TITLE OF INVENTION: STE20-RELATED PROTEIN KINASES
; FILE REFERENCE: 038602/0329
; CURRENT APPLICATION NUMBER: US/09/291,417D
; CURRENT FILING DATE: 1999-04-13
; PRIOR APPLICATION NUMBER: 60/081,784
; PRIOR FILING DATE: 1998-04-14

; NUMBER OF SEQ ID NOS: 155
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 31
; LENGTH: 1001
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: Synthetic
; OTHER INFORMATION: murine/human SULU3
US-09-688-188B-31

Query Match 9.2%; Score 248.5; DB 4; Length 1001;
Best Local Similarity 21.6%; Pred. No. 8.6e-09;
Matches 139; Conservative 117; Mismatches 197; Indels 191; Gaps 29;

QY 23 PAVEAE-----GPGSSQA-PRKPEGAQARTAGSALRDVSEELSRQLE 64
Db 324 PAVEAQEEBEEQDHGVRGTGTVNSVGSNQSIPTSMISASSQSSVNSLFDVDDKS-ELD 382
QY 65 DILSTCYVNNQO-----GPGEDG--AQGEPAEPEDAERKSTTVARNGEPEPTPVVYGEKE 118
Db 383 MMEGDHTVMSNSSVIHLKPEENYREGDP-----RT---RASDPQSPQVSRHKS 430
QY 119 PSKGDPTNTEIROSDVGRDRHRRPOEKKAKGLGKEIT-----LLMQTLNTLSTPEEKLA 174
Db 431 HYNRREHPATIRTASLV---TRQMEHEQDSSELREQMSGYKMRROHQKQMLTLENK 486
QY 175 ALCKKYA-----ELLEHRN--SOKMKLLQKQSOLVQE-----KDHLCGEHSK 217
Db 487 AEMDEHRLDKOLETORNNFAAEKELIKKHQAAMEKEAKVMSNEEKKFOQHIOAQQCK 546
QY 218 AVLARSKLESICRELQHRNRSKKEGVQR-----AREEEE---KRKEVTSHFQV--TLNDI 268
Db 547 EL--NSFLESQKREYKLRKEQLKEELNENQSTPKKEQEWLSKQENIQHFQABEEANLL 604
QY 269 QLOME-----QHN-----BRNSKLQENMELAEKLLKLEIQVYELRE 304
Db 605 RRQROYLELECRFRKRMMLGRENLEQDLVREELNKRQTKDLEHAMLRRQHESMOELEF 664
QY 305 EHIDKVFKHDLQOOLVDKALQQAQAEMLKEAEERHOREKDFLLKEAVESQRMCELMKQOE 364
Db 665 RHLNTIQK--MRCELL--RLQHQTELNTQLEYNKRERELRKHVMEVRQPKSLKSKE 719
QY 365 THLKOQL-----ALYTKPEEFONTLSKSEVFTTFQEMEKMTYKIKKLEKETTMYRSRW 420
Db 720 LQIKKQFQDTCKIQTRYKALRNHLE-----TTPKSEHKAVLKRLK--BEQTRKLAILA 772
QY 421 ESSNKALLEMAEKTVR-----DKELE 442
Db 773 EYDHSINEMLSQALRLDEQAECQVLMQLOQLELLNAYQSKIKNQAEACHDRER 832
QY 443 GLQVKI-----QRLEKLCRALQTERND-----LNKRVDLSAGGQGS-----L 480
Db 833 ELQRVSLRRALLEQKIEEEMALQNERTERIRSLERQARIEAFDESRLGFSNMVL 892
QY 481 TDSGPERPEPGAQAPSSPRVTEAPCYGAP--STEASQOTGP 522
Db 893 SNLSPE-----AFSHSYPGASGWSHNPTGGPGP 920

RESULT 13
US-09-688-188B-151
; Sequence 151 Application US/09688188B
; Patent No. 6656716
; GENERAL INFORMATION:
; APPLICANT: PLOWMAN, GREGORY
; APPLICANT: MARTINEZ, RICARDO
; APPLICANT: WHYTE, DAVID
; TITLE OF INVENTION: STE20-RELATED PROTEIN KINASES
; FILE REFERENCE: 038602/0328
; CURRENT APPLICATION NUMBER: US/09/688,188B
; CURRENT FILING DATE: 2000-10-16

; NUMBER OF SEQ ID NOS: 155
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 31
; LENGTH: 1001
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: Synthetic
; OTHER INFORMATION: murine/human SULU3
US-09-688-188B-31

Query Match 9.2%; Score 248.5; DB 4; Length 1001;
Best Local Similarity 21.6%; Pred. No. 8.6e-09;
Matches 139; Conservative 117; Mismatches 197; Indels 191; Gaps 29;

QY 23 PAVEAE-----GPGSSQA-PRKPEGAQARTAGSALRDVSEELSRQLE 64
Db 324 PAVEAQEEBEEQDHGVRGTGTVNSVGSNQSIPTSMISASSQSSVNSLFDVDDKS-ELD 382
QY 65 DILSTCYVNNQO-----GPGEDG--AQGEPAEPEDAERKSTTVARNGEPEPTPVVYGEKE 118
Db 383 MMEGDHTVMSNSSVIHLKPEENYREGDP-----RT---RASDPQSPQVSRHKS 430
QY 119 PSKGDPTNTEIROSDVGRDRHRRPOEKKAKGLGKEIT-----LLMQTLNTLSTPEEKLA 174
Db 431 HYNRREHPATIRTASLV---TRQMEHEQDSSELREQMSGYKMRROHQKQMLTLENK 486
QY 175 ALCKKYA-----ELLEHRN--SOKMKLLQKQSOLVQE-----KDHLCGEHSK 217
Db 487 AEMDEHRLDKOLETORNNFAAEKELIKKHQAAMEKEAKVMSNEEKKFOQHIOAQQCK 546
QY 218 AVLARSKLESICRELQHRNRSKKEGVQR-----AREEEE---KRKEVTSHFQV--TLNDI 268
Db 547 EL--NSFLESQKREYKLRKEQLKEELNENQSTPKKEQEWLSKQENIQHFQABEEANLL 604
QY 269 QLOME-----QHN-----BRNSKLQENMELAEKLLKLEIQVYELRE 304
Db 605 RRQROYLELECRFRKRMMLGRENLEQDLVREELNKRQTKDLEHAMLRRQHESMOELEF 664
QY 305 EHIDKVFKHDLQOOLVDKALQQAQAEMLKEAEERHOREKDFLLKEAVESQRMCELMKQOE 364
Db 665 RHLNTIQK--MRCELL--RLQHQTELNTQLEYNKRERELRKHVMEVRQPKSLKSKE 719
QY 365 THLKOQL-----ALYTKPEEFONTLSKSEVFTTFQEMEKMTYKIKKLEKETTMYRSRW 420
Db 720 LQIKKQFQDTCKIQTRYKALRNHLE-----TTPKSEHKAVLKRLK--BEQTRKLAILA 772
QY 421 ESSNKALLEMAEKTVR-----DKELE 442
Db 773 EYDHSINEMLSQALRLDEQAECQVLMQLOQLELLNAYQSKIKNQAEACHDRER 832
QY 443 GLQVKI-----QRLEKLCRALQTERND-----LNKRVDLSAGGQGS-----L 480
Db 833 ELQRVSLRRALLEQKIEEEMALQNERTERIRSLERQARIEAFDESRLGFSNMVL 892
QY 481 TDSGPERPEPGAQAPSSPRVTEAPCYGAP--STEASQOTGP 522
Db 893 SNLSPE-----AFSHSYPGASGWSHNPTGGPGP 920

RESULT 12
US-09-291-417D-31
; Sequence 31 Application US/09291417D
; Patent No. 6680170
; GENERAL INFORMATION:
; APPLICANT: PLOWMAN, GREGORY
; APPLICANT: MARTINEZ, RICARDO
; APPLICANT: WHYTE, DAVID
; TITLE OF INVENTION: STE20-RELATED PROTEIN KINASES
; FILE REFERENCE: 038602/0329
; CURRENT APPLICATION NUMBER: US/09/291,417D
; CURRENT FILING DATE: 1999-04-13
; PRIOR APPLICATION NUMBER: 60/081,784

; PRIOR APPLICATION NUMBER: 09/291,417
; PRIOR FILING DATE: 1998-04-14
; PRIOR APPLICATION NUMBER: 60/081,784
; PRIOR FILING DATE: 1998-04-14
; NUMBER OF SEQ ID NOS: 155
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 151
; LENGTH: 787
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-688-188B-151

Query Match 9.1%; Score 247; DB 4; Length 787;
Best Local Similarity 21.7%; Pred. No. 8.1e-09;
Matches 139; Conservative 117; Mismatches 206; Indels 178; Gaps 29;
QY 14 AQERPSQAAPAVEAEP-----GSSQA-PRKPEGAQAATAQSGALRDVSEELSRQLEDILS 68
DB 114 AQEEBEEQDHGVRGTGTVNSVGSNQSIPSMISASQSSVNSLFDVSDDKS-ELDMWEG 172
QY 69 TYCVNNQG---GPGEDG--AQGEPAEPEDAESRTYVARNGPEPTPVYVGEKPSKG 122
DB 173 DHTVNSNSVIHLKPEENYREGDP-----RT---RASDPQSPQVSRHSHYRN 220
QY 123 DNTTEIRQSDVGDHRRPQEKKAKGLGKEIT---LLMQTLNTLSTPEEKLAALCK 178
DB 221 REHFATIRTASLV---TRQMEHODSELREQMSGYKMRERQHQKQLMTLENKLAEMD 276
QY 179 KYA-----ELLEHRN--SOKQMKLLQKQSLVQE-----KDLRGEHSHKAVLA 221
DB 277 EHRLLDKLGTGRNFAAEMKLIKQQAAMEKAKVMSNEEKFFQOHIQAQKKEL-- 334
QY 222 RSKLESCLRELQRHNSLKEGVOR---AREEBE---KRKEVTSHFQV--TLNDIQLQM 272
DB 335 NSFLESQKEEYKLRKEQLKEELNENQSTPKKEQEWLSKOKENIQHFOABEENLLRRQR 394
QY 273 E-----OHN-----ERNKLRQENMELAEKLIKLIQOYELREEHID 308
DB 395 QYLEECRRFRKRMMLGRHNLQDLVBEELNKRQTKDLEHMLLRQHESMCELEFRHLN 454
QY 309 KVFKEHKLQOQLVDKLAQQAQEMLEKAEERHOREKDFLLKEAVESQRMCELMKQOETHLK 368
DB 455 TIQK---MRCELI--RLQHOTELTNQLEYNKRERELRRKHVMEVRQPKSLKSELOIK 509
QY 369 QOL---ALYTEKPEFQNTLSKSEVFTTFQKEMEMTKIKKLEKETMYRGRWESSN 424
DB 510 KQFQDTCKIQTRYKALRNHLE---TTPKSEHKAVALKRLK--EQOTRKLAILAEQYD 562
QY 425 KALLEMAEKTVR-----DKELEGLOV 446
DB 563 HSNEMLSQAIRLDEAQAECQVLMQLOQLELLNAYQSKIKWQAQAQHDRLRELEQ 622
QY 447 KI-----QRLEKLCRALQTERND-----LNKRVQDLSAGGGS-----LTDG 484
DB 623 RVSLRRALLEQKIEEEMALQNERTERIRSLRQAREIFAQDSMELGFSNMVLSNLS 682
QY 485 PRRPEGGAQAPSSPRVTEAPCYGPAP--STEASGQTGP 522
DB 683 PE-----AFSHSYFGASGWSHNPTGGFGP 706

RESULT 14
US-09-291-417D-151
; Sequence 151, Application US/09291417D
; Patent No. 6880170
; GENERAL INFORMATION:
; APPLICANT: PLOWMAN, GREGORY
; APPLICANT: MARTINEZ, RICARDO
; APPLICANT: WHYTE, DAVID
; TITLE OF INVENTION: STE20-RELATED PROTEIN KINASES
; FILE REFERENCE: 038602/0329
; CURRENT FILING DATE: US/09/291,417D
; CURRENT FILING DATE: 1999-04-13

; PRIOR APPLICATION NUMBER: 60/081,784
; PRIOR FILING DATE: 1998-04-14
; NUMBER OF SEQ ID NOS: 155
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 151
; LENGTH: 787
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-291-417D-151

Query Match 9.1%; Score 247; DB 4; Length 787;
Best Local Similarity 21.7%; Pred. No. 8.1e-09;
Matches 139; Conservative 117; Mismatches 206; Indels 178; Gaps 29;
QY 14 AQERPSQAAPAVEAEP-----GSSQA-PRKPEGAQAATAQSGALRDVSEELSRQLEDILS 68
DB 114 AQEEBEEQDHGVRGTGTVNSVGSNQSIPSMISASQSSVNSLFDVSDDKS-ELDMWEG 172
QY 69 TYCVNNQG---GPGEDG--AQGEPAEPEDAESRTYVARNGPEPTPVYVGEKPSKG 122
DB 173 DHTVNSNSVIHLKPEENYREGDP-----RT---RASDPQSPQVSRHSHYRN 220
QY 123 DNTTEIRQSDVGDHRRPQEKKAKGLGKEIT---LLMQTLNTLSTPEEKLAALCK 178
DB 221 REHFATIRTASLV---TRQMEHODSELREQMSGYKMRERQHQKQLMTLENKLAEMD 276
QY 179 KYA-----ELLEHRN--SOKQMKLLQKQSLVQE-----KDLRGEHSHKAVLA 221
DB 277 EHRLLDKLGTGRNFAAEMKLIKQQAAMEKAKVMSNEEKFFQOHIQAQKKEL-- 334
QY 222 RSKLESCLRELQRHNSLKEGVOR---AREEBE---KRKEVTSHFQV--TLNDIQLQM 272
DB 335 NSFLESQKEEYKLRKEQLKEELNENQSTPKKEQEWLSKOKENIQHFOABEENLLRRQR 394
QY 273 E-----OHN-----ERNKLRQENMELAEKLIKLIQOYELREEHID 308
DB 395 QYLEECRRFRKRMMLGRHNLQDLVBEELNKRQTKDLEHMLLRQHESMCELEFRHLN 454
QY 309 KVFKEHKLQOQLVDKLAQQAQEMLEKAEERHOREKDFLLKEAVESQRMCELMKQOETHLK 368
DB 455 TIQK---MRCELI--RLQHOTELTNQLEYNKRERELRRKHVMEVRQPKSLKSELOIK 509
QY 369 QOL---ALYTEKPEFQNTLSKSEVFTTFQKEMEMTKIKKLEKETMYRGRWESSN 424
DB 510 KQFQDTCKIQTRYKALRNHLE---TTPKSEHKAVALKRLK--EQOTRKLAILAEQYD 562
QY 425 KALLEMAEKTVR-----DKELEGLOV 446
DB 563 HSNEMLSQAIRLDEAQAECQVLMQLOQLELLNAYQSKIKWQAQAQHDRLRELEQ 622
QY 447 KI-----QRLEKLCRALQTERND-----LNKRVQDLSAGGGS-----LTDG 484
DB 623 RVSLRRALLEQKIEEEMALQNERTERIRSLRQAREIFAQDSMELGFSNMVLSNLS 682
QY 485 PRRPEGGAQAPSSPRVTEAPCYGPAP--STEASGQTGP 522
DB 683 PE-----AFSHSYFGASGWSHNPTGGFGP 706

RESULT 15
US-09-060-410-2
; Sequence 2, Application US/09060410
; Patent No. 6155461
; GENERAL INFORMATION:
; APPLICANT: Cobb, Melanie
; APPLICANT: Hutchinson, Michele
; APPLICANT: Chen, Zhu
; APPLICANT: Bertram, Kevin
; TITLE OF INVENTION: TAO PROTEIN KINASES AND METHODS OF USE
; NUMBER OF SEQUENCES: 26
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SEED and BERRY LLP

STREET: 6300 Columbia Center, 701 Fifth Avenue
City: Seattle
STATE: Washington
COUNTRY: USA
ZIP: 98104
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/060,410
FILING DATE: 14-APR-1998
CLASSIFICATION:
ATTORNEY/AGENT INFORMATION:
NAME: Maki, David J.
REGISTRATION NUMBER: 31,392
REFERENCE/DOCKET NUMBER: 860098.421
TELECOMMUNICATION INFORMATION:
TELEPHONE: (206) 622-4900
TELEFAX: (206) 682-6031
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 1001 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-09-060-410-2

Query Match 9.1%; Score 247; DB 3; Length 1001;
Best Local Similarity 21.9%; Pred. No. 1.1e-08;
Matches 142; Conservative 118; Mismatches 212; Indels 176; Gaps 30;

QY 23 PAVEAE-----GPGSSQA-PEKPEGAQAARTAGSALRDVSEELSRQLE 64
DB 324 PAVEAESEEBQDHGGGRTGTWNSVGNQSIIPMSISASSQSSVNSLPDASDDKS-ELD 382
QY 65 DILSTYCVNNQ-----GPGEDGAQGEPAEPDAEKSRITYVARNGEPEPTPVVYGEKPS 120
DB 383 MMEGDTWMSNVHLKPEENYQ-EGDPR-----TRASAPQSPQVSRHSHY 432
QY 121 KGPNTTEIRQSDVGDHRRRPOBKAKGLGKEIT-----LLMQTLNTLSTPEKLAAL 176
DB 433 RNEHFATIRTSALV-----TRQMEHQDSLEQMSGYKRMRRQKQMLTLENKLKAE 488
QY 177 CKKYA----ELLEHRN--SOKMKLLQKQSLVQ-----KDLRGEHSHY 219
DB 722 IRKQFQDTCKIQTRQYKAUNHLL-----TPKSEHKAVLKKL--SEQTKLAILAQ 774
QY 423 SNKALLEMAEKTVR-----DKELEG 444
DB 775 YDHSINEMLSQALRLDEAQAEQVLMQLOQOELELLNAYQSKIWKQAEQAHDREL 834
QY 445 QVKI-----QRLKLCALQTERND-----LNRVQDLSGGGGS-----LTD 482
DB 835 EQRVSRALLQKIEEMALQNERTHIRSLRQAEIEAFDSBMSMLGFSNNVLN 894

QY 483 SGPERRPEG-EGQAQPS-SPRVTRAP--CYPGAPSTEASG---QTGPQ 523
DB 895 LSPFAFSHYFGASSWSHNPCTGCGPHWGHMGGTPOAWGHMGGGPQ 942
RESULT 15
US-09-723-458-2
Sequence 2, Application US/09723458
Patent No. 6586242
GENERAL INFORMATION:
APPLICANT: Cobb, Melanie
Hutchinson, Michele
Chen, Zhu
Berman, Kevin
TITLE OF INVENTION: TAO PROTEIN KINASES AND METHODS OF USE THEREFOR
NUMBER OF SEQUENCES: 26
CORRESPONDENCE ADDRESS:
ADDRESSEE: SEED and BERRY LLP
STREET: 6300 Columbia Center, 701 Fifth Avenue
City: Seattle
STATE: Washington
COUNTRY: USA
ZIP: 98104
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/723,458
FILING DATE: 27-NO. 6586242-2000
CLASSIFICATION: <unknown>
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/09/060,410
FILING DATE: <unknown>
ATTORNEY/AGENT INFORMATION:
NAME: Maki, David J.
REGISTRATION NUMBER: 31,392
REFERENCE/DOCKET NUMBER: 860098.421
TELECOMMUNICATION INFORMATION:
TELEPHONE: (206) 622-4900
TELEFAX: (206) 682-6031
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 1001 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
SEQUENCE DESCRIPTION: SEQ ID NO: 2:
US-09-723-458-2

Query Match 9.1%; Score 247; DB 4; Length 1001;
Best Local Similarity 21.9%; Pred. No. 1.1e-08;
Matches 142; Conservative 118; Mismatches 212; Indels 176; Gaps 30;

QY 23 PAVEAE-----GPGSSQA-PRPEGAQAARTAGSALRDVSEELSRQLE 64
DB 324 PAVEAESEEBQDHGGGRTGTWNSVGNQSIIPMSISASSQSSVNSLPDASDDKS-ELD 382
QY 65 DILSTYCVNNQ-----GPGEDGAQGEPAEPDAEKSRITYVARNGEPEPTPVVYGEKPS 120
DB 383 MMEGDTWMSNVHLKPEENYQ-EGDPR-----TRASAPQSPQVSRHSHY 432
QY 121 KGPNTTEIRQSDVGDHRRRPOBKAKGLGKEIT-----LLMQTLNTLSTPEKLAAL 176
DB 433 RNEHFATIRTSALV-----TRQMEHQDSLEQMSGYKRMRRQKQMLTLENKLKAE 488
QY 177 CKKYA----ELLEHRN--SOKMKLLQKQSLVQ-----KDLRGEHSHY 219
DB 489 MDEHRLDKOLETRQNFPAEMKLIKQASMEKAKYMANBEKKQOHIQAKKEL 548
QY 220 LARSKLESCLRELQRNHSKEEGVQR-----ARBEES---KXKEVTSHFQV--TLNDIQL 270

Db 549 --NSFLESQRYKRLKKEQLKEELNENOSTPKKBEQOEWLSKQENIQHPQAEERANLRR 606
QY 271 QME-----QHN-----ERNKLRQENMELAEKLLIQEYELREBH 306
Db 607 ORQYLECERFRKRMMLGRHNLQDVLVEELNKRTQKDLHMLRQHESQOELFRH 666
QY 307 IDKVFHKDLOQLVDKIQOQOEWLKEAEERHOREKOPFLLEAVESQMCMLMQOETH 366
Db 667 LNTIQK---MRCELI---RLQHOTELTNQLEYNKRERELRRHVMVEVROQPKSLKSLQ 721
QY 367 LKQOL---ALVTEKPEEPONTLSKSSEVFTTFKQEMERMTKKIKLEKETTMYRSWES 422
Db 722 IKKQDQTKCIQTRQYKALRNHLZ-----TPPKSEHKAVLRKX--BEQTKLAILAEQ 774
QY 423 SNKALLEMAEBKTVR-----DKELEGE 444
Db 775 YDHSINEMLSQALRLDEAEQAVLQKQOELNAYQSKIMQOAEHQHRELREL 834
QY 445 QVKI-----ORLEKLCRALQTERND-----LNKRVODLSAGCGS-----LTD 482
Db 835 EQRVSRLRALLQKIEEMALQNERIRISLLERQAREIFAQSESRVLGFSNMVLN 894
QY 483 SOPERREPG-PGAQAPS-SPRVTEAP--CYPGAPSTEASG---QTPCP 523
Db 895 LSPFAFHSYPGASSWSHNPTGSGPHMGHPMGGTPQAWGHPMQGPP 942

RESULT 17

US-09-595-684B-31

; Sequence 31, Application US/09595684B

; Patent No. 6544766

; GENERAL INFORMATION:

; APPLICANT: Beraud, Christophe

; APPLICANT: Ohashi, Cara

; APPLICANT: Sakowicz, Roman

; APPLICANT: Vaisberg, Eugeni

; APPLICANT: Wood, Kenneth

; APPLICANT: Yu, Ming

; TITLE OF INVENTION: Human kinesins and methods of producing

; FILE REFERENCE: cytop036

; CURRENT APPLICATION NUMBER: US/09/595,684B

; PRIOR FILING DATE: 2002-06-24

; PRIOR APPLICATION NUMBER: 09/295,612

; NUMBER OF SEQ ID NOS: 105

; SOFTWARE: FastSeq for Windows Version 4.0

; SEQ ID NO 31

; LENGTH: 2662

; TYPE: PRT

; ORGANISM: Human

US-09-595-684B-31

Query Match

Best Local Similarity 8.9%; Score 240; DB 4; Length 2662;

Matches 130; Conservative 101; Mismatches 176; Indels 238; Gaps 25;

QY 55 VSELSROLEDILSTYCVDDNNGGPGEDGAQ-----EPAEPEDAERSYVA----- 102

Db 986 ISEVSRNLH-----MEENTGETKDFQQRWVGIDKKQDLKAKNTQTLTADVADNEII 1038

QY 103 -----RNGEPFTPVYGEKPSKGD-----PNTTEIR-QSDEVGDEDH 140

Db 1039 EQRKIFSLIOBKNELOQLMSVIAEKQDLTKENIEMTIENQEBELRLIGDLKQOE 1098

QY 141 RPPQEKKA-KGLGKEITLLMTLSTPBEKL----- 173

Db 1099 IVAQEKNAIKKEG-----LSRTCDRLAEVEKLEKESQQLQEQQLNQEEMSEMQ 1154

QY 174 -----AALCKYAELEBEHNSQOMKLLQKQSOLVQS 207

Db 1155 KINEIENLKNELKNKELTLEHMETERLELAQKLENYEVKSIYKRVKQLQKSPETE 1214

QY 208 KDHLRG-----EHSKAV--LARSKLESCL-----RELQRHN 236
Db 1215 RDHLRGYRETEATGLQTKELKIAHILKEHQETIDELRSVSEKTAQIINTQDLKSH 1274
QY 237 RSLKEEGVQAREREEKEKEV--TSHFOVTLNDIQLQMEQHNERNS-----KLROENMELA 290
Db 1275 TKLQEE-IPVLHBEQELLPNVKVSETQETNNLELLTEQSTTKDSTTLARIEMERLNL 1333
QY 291 ERLKKLIQYELREBEHIDKV-----FKHKLQOQLVD--AKLQQAQ----- 329
Db 1334 EKFBSSQBEIKSLTKERDNLTKIKEALEVKGHDQLKEHIRETLAKIQESQSQEQSLANKE 1393
QY 330 -----EML-----KEAERHORBKDFLLKEAVESQMCML 359
Db 1394 KONETTKIVSEMOFKPKDSALLRIEIMLGLSKLEQESHDMKS--VAKEXDDQLQLOEV 1452
QY 360 MKQOETHLKKOL-----ALYTEKPEEFO-----NTLSKSESEVFTTFKQ 397
Db 1453 LQSSDQLKENIKVAKHLSLEELKVAHCLKEQSETINELRVNLSEKETEISTIQK 1512
QY 398 ---EMEKWKIKKL-EKETM---YRSWESSKALLEMAEKTVDKLEGLQVKI-- 448
Db 1513 LEANDKLQNKIQBIYEKEBOLAKOISEVOENVNELKQFKHRAKAKDSALQSIESKMLE 1572
QY 449 -----ORLEKLCRALQTERNDLNKRVODLSA 474
Db 1573 LTNRLQSBQEBIQIMIKEXEMKRVQELQIERDQLKENTYKIVA 1617

RESULT 18

US-08-056-200-94

; Sequence 94, Application US/08056200

; Patent No. 5616500

; GENERAL INFORMATION:

; APPLICANT: Steinert, Peter M.

; APPLICANT: Lee, Seung-Chul

; APPLICANT: Kim, In-Gyu

; APPLICANT: Chung, Soo-Il

; APPLICANT: Park, Sang-Chul

; TITLE OF INVENTION: Trichohyalin and Transglutaminase-3 and

; TITLE OF INVENTION: Methods of Using Same

; NUMBER OF SEQUENCES: 117

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Knobbe, Martens, Olson & Bear

; STREET: 620 Newport Center Drive, Sixteenth Floor

; CITY: Newport Beach

; STATE: CA

; COUNTRY: U.S.A.

; ZIP: 92660

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; COMPUTER: IBM PC compatible

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: Patent In Release #1.0, Version #1.25

; CURRENT APPLICATION DATA: US/08/056,200

; FILING DATE: 30-APR-1993

; CLASSIFICATION: 435

; ATTORNEY/AGENT INFORMATION:

; NAME: Pedrick, Michael P.

; REGISTRATION NUMBER: 36,799

; REFERENCE/DOCKET NUMBER: NIH054.001A

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: (714) 760-0404

; TELEFAX: (714) 760-9502

; INFORMATION FOR SEQ ID NO: 94:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 1898 amino acids

; TYPE: amino acid

; TOPOLOGY: linear

; MOLECULE TYPE: protein

US-08-056-200-94

;/ TITLE OF INVENTION: Graft Rejection
;/ FILE REFERENCE: UCSF-090
;/ CURRENT APPLICATION NUMBER: US/09/310,187A
;/ CURRENT FILING DATE: 1999-05-12
;/ NUMBER OF SEQ ID NOS: 3
;/ SOFTWARE: FastSeq for Windows Version 4.0
;/ SEQ ID NO 1
;/ LENGTH: 1939
;/ TYPE: PRT
;/ ORGANISM: Homo sapiens
;/ US-09-310-187A-1

Query Match 8.8%; Score 236.5; DB 4; Length 1939;
Best Local Similarity 24.4%; Pred. No. 1.3e-07;
Matches 146; Conservative 81; Mismatches 196; Indels 175; Gaps 25;

Qy 12 EGACERPSQAAPVAEAGPGSSQAPKPEGAQARTAGSALRDVSELSQLEDILSTYC 71
Db 1037 EGSLEQEKVMDLE-----RAKRLGDLKLTQES--IMDLND-KLQLEEKLKKE 1086

Qy 72 VDNNGGPGDGAQGEPAEPEDA-----EKSRTYVAR-----NGEPPTPVVYGEKEP 119
Db 1087 FDIQ-----QNSKIEDEQALALQLOKLEKQARIEELEEAEARTAKVLEKL 1138

Qy 120 SKGDPNTEIROS-DEVG-----DRHRRPQKKKAKGLKGRITLLMOTLNTLSTPEE 171
Db 1139 SDLSRELEISERLEAGGATSQVIEENMKKRAEFQKWRDLLEATL-----QHEA 1189

Qy 172 KLAALCKKY-----ABILLEHRNSQOMKLLQKOSQOLVQEKDHLRGHSHKAVARSKLES 227
Db 1190 TAAALRKHADSVAEIGEIQDNLQKVKQKLEKSEFPLELDVTSNMEOIIRKAKANLEK 1249

Qy 228 LCR-----BLQRNRSLEKBEGVQVOR-----EKSRTYVAR-----NGEPPTPVVYGEKEP 248
Db 1250 VSRTELDQANVYKLEEAQRSLNDFTTQAKLQTENGELARQLEKEALISQITRGKLS 1309

Qy 249 -----ESEEKKEVTSH-FQVTLNDIQLOMEQHNER-----NSX 281
Db 1310 YTOQMEDLKQLEBEGKAKNALAHALQSAHDCDLLREQYEEETEAKAEIQRVLSKANSE 1369

Qy 282 LR-----QENMELAEKLLIEQYELREEHIDKV-----FKHK---DLQO 318
Db 1370 VAQWRTKYETDAIQRTTELEBAKKAQRLQDAEEAVEAVNAKCSSLEKTKHRLQNEIED 1429

Qy 319 QLVDAKLOQAQEMLKEAERHOREKDFLLKEAVESQRMCELMKQOETHLQKQALALYTEKF 378
Db 1430 LMVDVERSNA---AAALDKKQNFDKILAE--WKQKYEESQGELESSQKEARSLSSTELP 1484

Qy 379 BEFQNTLSKSSSEVFTTFKQEMKMTKKI-----KLEKETTMVR 417
Db 1485 -KLKAYEESLEHLETFKRNKNLOEBISDLTEQLEGGGKNNVHELEKVRKQLEVEKLELQ 1543

Qy 418 SRWESSNKALLEAEKTVRDKELEGQVKIQRLKLC---RALQTERDNLKRVQD 471
Db 1544 SALEEA-EASLEHEEGKILR-AQLEFNQIKABIERKLAKEDEMEQAKRM--HQRVVD 1597

RESULT 21
US-08-938-105-3
; Sequence 3, Application US/08938105
; Patent No. 6353151
; GENERAL INFORMATION:
; APPLICANT: Weinwand, Leslie A.
; APPLICANT: Vikstrom, Karen L.
; TITLE OF INVENTION: TRANSGENIC MODEL FOR HEART FAILURE
; NUMBER OF SEQUENCES: 3
; CORRESPONDENCE ADDRESS:
; ADDRESS: Sheridan Ross P.C.
; STREET: 1700 Lincoln St., Suite 3500
; CITY: Denver
; STATE: CO
; COUNTRY: U.S.A.
; ZIP: 80203

;/ COMPUTER READABLE FORM:
;/ MEDIUM TYPE: Floppy disk
;/ COMPUTER: IBM PC compatible
;/ OPERATING SYSTEM: PC-DOS/MS-DOS
;/ SOFTWARE: PatentIn Release #1.0, Version #1.30
;/ CURRENT APPLICATION DATA:
;/ APPLICATION NUMBER: US/08/938,105
;/ FILING DATE:
;/ CLASSIFICATION:
;/ ATTORNEY/AGENT INFORMATION:
;/ NAME: Crook, Wainell M.
;/ REGISTRATION NUMBER: 31,071
;/ REFERENCE/DOCKET NUMBER: 3595-4
;/ TELECOMMUNICATION INFORMATION:
;/ TELEPHONE: (303) 863-9700
;/ TELEFAX: (303) 863-0223
;/ INFORMATION FOR SEQ ID NO: 3:
;/ SEQUENCE CHARACTERISTICS:
;/ LENGTH: 1886 amino acids
;/ TYPE: amino acid
;/ TOPOLOGY: linear
;/ MOLECULE TYPE: protein
;/ US-08-938-105-3

Query Match 8.6%; Score 231.5; DB 4; Length 1886;
Best Local Similarity 24.4%; Pred. No. 2.7e-07;
Matches 146; Conservative 80; Mismatches 197; Indels 175; Gaps 25;

Qy 12 EGACERPSQAAPVAEAGPGSSQAPKPEGAQARTAGSALRDVSELSQLEDILSTYC 71
Db 984 EGSLEQEKVMDLE-----RAKRLGDLKLTQES--IMDLND-KLQLEEKLKKE 1033

Qy 72 VDNNGGPGDGAQGEPAEPEDA-----EKSRTYVAR-----NGEPPTPVVYGEKEP 119
Db 1034 FDIQ-----QNSKIEDEQALALQLOKLEKQARIEELEEAEARTAKVLEKL 1085

Qy 120 SKGDPNTEIROS-DEVG-----DRHRRPQKKKAKGLKGRITLLMOTLNTLSTPEE 171
Db 1086 SDLSRELEISERLEAGGATSQVIEENMKKRAEFQKWRDLLEATL-----QHEA 1136

Qy 172 KLAALCKKY-----ABILLEHRNSQOMKLLQKOSQOLVQEKDHLRGHSHKAVARSKLES 227
Db 1137 TAAALRKHADSVAEIGEIQDNLQKVKQKLEKSEFPLELDVTSNMEOIIRKAKANLEK 1196

Qy 228 LCR-----BLQRNRSLEKBEGVQVOR-----EKSRTYVAR-----NGEPPTPVVYGEKEP 248
Db 1197 VSRTELDQANVYKLEEAQRSLNDFTTQAKLQTENGELARQLEKEALISQITRGKLS 1256

Qy 249 -----ESEEKKEVTSH-FQVTLNDIQLOMEQHNER-----NSX 281
Db 1257 YTOQMEDLKQLEBEGKAKNALAHALQSAHDCDLLREQYEEEMEAKAEIQRVLSKANSE 1316

Qy 282 LR-----QENMELAEKLLIEQYELREEHIDKV-----FKHK---DLQO 318
Db 1317 VAQWRTKYETDAIQRTTELEBAKKAQRLQDAEEAVEAVNAKCSSLEKTKHRLQNEIED 1376

Qy 319 QLVDAKLOQAQEMLKEAERHOREKDFLLKEAVESQRMCELMKQOETHLQKQALALYTEKF 378
Db 1377 LMVDVERSNA---AAALDKKQNFDKILAE--WKQKYEESQGELESSQKEARSLSSTELP 1431

Qy 379 BEFQNTLSKSSSEVFTTFKQEMKMTKKI-----KLEKETTMVR 417
Db 1432 -KLKAYEESLEHLETFKRNKNLOEBISDLTEQLEGGGKNNVHELEKVRKQLEVEKLELQ 1490

Qy 418 SRWESSNKALLEAEKTVRDKELEGQVKIQRLKLC---RALQTERDNLKRVQD 471
Db 1491 SALEEA-EASLEHEEGKILR-AQLEFNQIKABIERKLAKEDEMEQAKRM--RYVD 1544

RESULT 22
US-09-914-259-11
; Sequence 11, Application US/09914259
; Patent No. 6495336

GENERAL INFORMATION:
; APPLICANT: Ito, Masaaki
; APPLICANT: Takahashi, No. 5906819uaki
; TITLE OF INVENTION: RHO TARGET PROTEIN RHO-KINASE
; NUMBER OF SEQUENCES: 16
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Foley & Lardner
; STREET: 3000 K Street, N.W., Suite 500
; CITY: Washington
; STATE: D.C.
; COUNTRY: USA
; ZIP: 20007-5109
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/685,576
; FILING DATE: 24-JUL-1996
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: JP 7-325129
; FILING DATE: 20-NOV-1995
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: JP 8-17150
; FILING DATE: 05-JAN-1996
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: JP 8-131206
; FILING DATE: 26-APR-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: Bent, Stephen A.
; REGISTRATION NUMBER: 29,768
; REFERENCE/DOCKET NUMBER: 16887/843
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202)672-5300
; TELEFAX: (202)672-5399
; TELEX: 904136
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1388 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-685-576-1

Query Match 8.4%; Score 228; DB 2; Length 1388;
Best Local Similarity 21.0%; Pred. No. 3.2e-07;
Matches 130; Conservative 83; Mismatches 177; Indels 228; Gaps 20;

QY 91 PRDAEKSR-----TYVARNGEPE-----PTPW--- 113
DB 316 PDAAETSKAKNLCAPLTDREVRIGRNVGVEIKQHPFKNDQWWDNIRETAAVPVPEL 375
QY 114 -----YGEKEPSKGDPTTEEI-----ROSDEVG 136
DB 376 SSDIDSSNFDIEDDDGDVETFPPIKAFVGNQLPFITGYRENLILLSDSKCKENDSIQ 435
QY 137 DRDHRPQ-----KKAKAGLGRKLTLLMQTL 163
DB 436 SRNEESQBIQKXLYTLBEHLSTEIOAKEELSQCKSVNTRLEKVAKELEBEITLRKQVE 495
QY 164 NTLSTPEEKLAALCKYAE-----LLEHRNSOKMK 195
DB 496 STLRQLEREKALLQHKNAEYQKADHEADKKNLENDVNSLKDQLDLKKNQNSQISTE 555
QY 196 LLOKQSQVLQVQKHRLRGEHSHKAVLAR-----SKLESICHELORHNSLKEEGVQ 245
DB 556 KYNQLQRQDQETNALLRTESDTAARLKTQAEBSKQIQOLESNNRDLQKNCILLETAKL 615
QY 246 RARE-----EEKREKVTSHFQVTLNDIQ-----LQMEQHN 276
DB 616 LEKEPINQSVLESERRDR--THGSEIINDLQGRISGLEEDVNGKILLAKVELEKRLQ 673

GENERAL INFORMATION:
; APPLICANT: Makowski, Lee
; APPLICANT: Hyman, Paul
; APPLICANT: Williams, Mark
; TITLE OF INVENTION: STAGED ASSEMBLY OF NANOSTRUCTURES
; FILE REFERENCE: 8471-010-999
; CURRENT APPLICATION NUMBER: US/09/914,259
; CURRENT FILING DATE: 2006-11-21
; NUMBER OF SEQ ID NOS: 180
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 11
; LENGTH: 3878
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-914-259-11

Query Match 8.5%; Score 230.5; DB 4; Length 3878;
Best Local Similarity 18.4%; Pred. No. 7.7e-07;
Matches 119; Conservative 110; Mismatches 195; Indels 223; Gaps 23;

QY 31 GSSQAPRKPGAQARTASGALRDVSELSRQLDILSTYCVNNO-----GGPGEDG 83
DB 161 GAQSPDTHLEMMSELA--GKQHEI--BELARELEEMVTVTGTGLQQLQSFRAAIKORDG 217
QY 84 -----AQGPAEPEDAERKSTVYVARNGEPEPTPVVYGEKEPSKGDPTNT----- 126
DB 218 IITOLTANLQAREKEDTWREFLELTQSQKLIQFQQLQASSETLNSHTSSTAADLLQ 277
QY 127 ---EEIQSDVGDGRDHRROEKKKANGKLGKBTLLMOTLNTLSTPEKLAALCKK----- 179
DB 278 AKQIILTHQOQLQEDQKLLDYQKKEDFTVQISFLQEKIKVYEMEQDKVENSKEEIQ 337
QY 180 -----YAELEHRNSOKM-----KLOKQSOLVQKDKHLRGEHSHKAVLAR 223
DB 338 EKETIIEELNTKIIEEKKTILELKLKLTADKLLGELQSQIVQKQKQIK--NMKLELTNS 395
QY 224 K-----LESICRELQRNR--SLKEGVQARAREBEKRYE----- 256
DB 396 KQKERSERIKQLMGTVBELQKXNHRDSQFETDIVQRMQETQKLEQLRAELDEMYGQ 455
QY 257 -----VTSH-----FOYTLNDIOLQME 273
DB 456 QIVQMKQELIRQMAQMEKTRKGENALRSYSNITVNEQIKLMVAINELIKLQ 515
QY 274 QHNRNSKLR-----QENNELAERLKLQTEQVELREHIDKVFK-----HK 314
DB 516 DTNSQKELKBEGLILBEKCALQRLQEDLVEELSPSRQIQARQTIARQESKLINEAHK 575
QY 315 DLQO-QLVADKLQQAQEMLKEABERH-----OREKDFLLKEAVESQRM--- 356
DB 576 SLSTVEDLKAEIVSASSRKELELKHAEVTVNYKIKLEMLEKEKNAVDNRMAESQEALE 635
QY 357 -----CELMKQOE-----TLKQQL-----ALYTEKF 378
DB 636 RLRTQLLFSHEBELSKLEDLEIEHRINIEKLNKONLGIHYKQIDGLQNMESQKIETMQF 695
QY 379 BEFQNTLSKSEVF-----TTFKQEMKMTKKIKLEKETTMYRSRWSSNKAL 427
DB 696 EK-DNLITKQNLILSILKLDQOOSLVNSKSEMTLQINELQKEIEILR----- 744
QY 428 LEMAEKTVRDKEGLEQVQKIQRLKCLRALQTERNDLNRVQDLA 474
DB 745 -QEEKRGTLQEQVQELQKLTLELEK---QMKKENDLQKFAQLEA 787

RESULT 23
US-08-685-576-1
; Sequence 1, Application US/08685576
; Patent No. 5906819
; GENERAL INFORMATION:
; APPLICANT: Kaibuchi, Kozi
; APPLICANT: Iwamatsu, Akihiro
; APPLICANT: Nakano, Takeshi

Db 366 KLTDLSQCORNAESARCSELEQKIK---EKEPEQBELSRQORSFQTLDDQECIQMKARLT 422
QY 206 QEKDHLRGHSAVLARSKLSLSCRELQHRNLSKEGVQARBEERKEKKEVTSHFQVTL 265
Db 423 QELQAKMHNVLQALDKLTSVKQLENNLEEFKQ---KLCSAQAFQAS-----QIKE 474
QY 266 NDIQLOMEQHNRNSKLRQENNELA-----ERLKKLIEQYELREBHIDKVPFKHDLQOQ 319
Db 475 NEILRSMEEMKKNLLKSHSQKAREVCHLEAEKLNKIQCLNOSQNFPAEEMKAKNTSQE 534
QY 320 LVDAKLQQAQEMLKEAEERHQRKDFLLKAEAVESQRM---ELMKQOETHLKQQLALYTEK 377
Db 535 TM---LRDLQEKINQOENSLTLEKLAVALDEKQDCSDLLKKEHHI-----581
QY 378 FFEQNTLSKSSEVFTTFKQEMEKMTKKIKLEKTTMYRSWESSNKALLEMABEKTVR 437
Db 582 -EQLNDKLSKTEKESKALLSALSELKKEYEELKEBKTILF-SCKWSENEKLL-----TQM 633
QY 438 DKELEGLOVKIQLEKLCALQTERNDLNKRVQDL 472
Db 634 ESEKENLQSKINHLETCLTKTQOIKSHYNERVRTL 668

RESULT 28
PCT-US95-16216-1
; Sequence 1, Application PC/TUS9516216
; GENERAL INFORMATION:
; APPLICANT: Yen, Timothy J.
; APPLICANT: Rattner, Jerome B.
; TITLE OF INVENTION: Nucleic Acid Encoding a Transiently
; TITLE OF INVENTION: Expressed Kinetochores Protein, and Methods of Use
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Darr, Dorfman, Herrell and Skillman
; STREET: 1601 Market Street Suite 720
; CITY: Philadelphia
; STATE: PA
; COUNTRY: USA
; ZIP: 19103-2307
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US95/16216
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/353,700
; FILING DATE: 09-DEC-1995
; ATTORNEY/AGENT INFORMATION:
; NAME: Reed, Janet B.
; REGISTRATION NUMBER: 36,252
; TELEPHONE: (215) 563-4100
; TELEFAX: (215) 563-4044
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 3248 amino acids
; TYPE: amino acid
; STRANDEDNESS: not relevant
; TOPOLOGY: not relevant
; MOLECULE TYPE: protein
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
PCT-US95-16216-1

Query Match 8.4%; Score 226.5; DB 5; Length 3248;
Best Local Similarity 22.4%; Pred. No. 1.2e-06;
Matches 102; Conservative 91; Mismatches 191; Indels 71; Gaps 14;
QY 31 GSSQAPKPEGAQTAAQSALRDVSELSRQLEDILSTVCVDDNNQGGPCDGAQCEPAE 90

Db 272 GNSSPHLLDQK---AQNLNRKINLEURL-----QGH--E 305
QY 91 PEDAEKSTYVARNGEPEPTPVVTGEKEPSKGDPTBEIRQSDVEGDRHRRPOEKKKAK 150
Db 306 KEMGQVNFQELQLEKAKVELLEKVKLNKCRDELVRTTAQYDQASTKYTALEQKXK 365
QY 151 GLGKEITLMTLNTL-STPEEKALALCKKVAELLEHRNSQKQKLLQKK-----QSQIV 205
Db 366 KLTDLSQCORNAESARCSELEQKIK---EKEPEQBELSRQORSFQTLDDQECIQMKARLT 422
QY 206 QEKDHLRGHSAVLARSKLSLSCRELQHRNLSKEGVQARBEERKEKKEVTSHFQVTL 265
Db 423 QELQAKMHNVLQALDKLTSVKQLENNLEEFKQ---KLCSAQAFQAS-----QIKE 474
QY 266 NDIQLOMEQHNRNSKLRQENNELA-----ERLKKLIEQYELREBHIDKVPFKHDLQOQ 319
Db 475 NEILRSMEEMKKNLLKSHSQKAREVCHLEAEKLNKIQCLNOSQNFPAEEMKAKNTSQE 534
QY 320 LVDAKLQQAQEMLKEAEERHQRKDFLLKAEAVESQRM---ELMKQOETHLKQQLALYTEK 377
Db 535 TM---LRDLQEKINQOENSLTLEKLAVALDEKQDCSDLLKKEHHI-----581
QY 378 FFEQNTLSKSSEVFTTFKQEMEKMTKKIKLEKTTMYRSWESSNKALLEMABEKTVR 437
Db 582 -EQLNDKLSKTEKESKALLSALSELKKEYEELKEBKTILF-SCKWSENEKLL-----TQM 633
QY 438 DKELEGLOVKIQLEKLCALQTERNDLNKRVQDL 472
Db 634 ESEKENLQSKINHLETCLTKTQOIKSHYNERVRTL 668

RESULT 29
US-08-195-487-4
; Sequence 4, Application US/08195487
; Patent No. 5783403
; GENERAL INFORMATION:
; APPLICANT: TOURKATLY, GARY
; APPLICANT: LIDGARD, GRAHAM P
; TITLE OF INVENTION: NOVEL MALIGNANT CELL TYPE MARKERS OF THE
; TITLE OF INVENTION: INTERIOR NUCLEAR MATRIX
; NUMBER OF SEQUENCES: 6
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: TESTA HURWITZ & THIBEAULT
; STREET: 53 STATE STREET
; CITY: BOSTON
; STATE: MA
; COUNTRY: USA
; ZIP: 02109
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/195,487
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/07/901,701
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: FITCHER ESQ, EDMUND R
; REGISTRATION NUMBER: 27,829
; REFERENCE/DOCKET NUMBER: MTP-013
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 617/248-7000
; TELEFAX: 617/248-7100
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2101 amino acids
; TYPE: amino acid
; TOPOLOGY: linear

MOLECULE TYPE: protein
US-08-195-487-4

Query Match 8.3%; Score 223; DB 1; Length 2101;
Best Local Similarity 22.1%; Pred. No. 1.2e-06;
Matches 111; Conservative 91; Mismatches 229; Indels 72; Gaps 17;

42 AQARTAGSALRDVSELSQLEEDILSTYCVNNQGGPGBD-GAOGEPAPEDAKSRIV 100
Db 1296 ASGREAEK--QRVASENLRQEL-----TSQAEAEELQELKAWQKFFQKEQAL 1343

101 VARGNEPEPTPVVYGEKPSKGDPTNTEIRQSDVEGDRDHR--POEKKAKGLGKEITL 158
Db 1344 STLQLEHTSTQALVSELLPAK--HLCQQLAQEAQAAAEKRRHELEQSKQAAGLRAELLR 1401

159 LMQTLNLTSTPEEKLAALCKKYAELEHRSNOKMLKQKQOLVQEKDHLRGHESKA 218
Db 1402 AQRELGLIIFERQVAEODRTAQQLRAEKASVAEQLSMLKKAHGLLAEN---RGLGERA 1458

219 VLARSKLESLEL--QRNRSLSKEGVQRAEEREEKKEVTSHPQVTLNDIOLQMSQHNE 277
Db 1459 NLGROFLV---ELDOAREKYVQELAAVRA--DAETRLAEVQREBAQSTARELEVMTAKYEG 1514

278 RNSKLQENMELAEKLLIEQVLEHREHIDKVFHKDLQOQLVDKLOQAQEMLKEABE 337
Db 1515 AKVLEERQRFQERQKLTQVE-----ELSKKLADS-----DQASK 1552

338 RHQREKDFLKEAVESQRMCELMKQETHLKOQALYTEKPEEFQNTLSKSSSEVFTTFKQ 397
Db 1553 VQOQKLKAVQAQGGESQQAQRFQALNELQALSKQEAHVKLQMEKAKTHYDAKKQ 1612

398 EMEKMTKIKKLEKETTMYRSWESSNKALLEWABEKTVRDKELEGLOVKIQLEKLCRA 457
Db 1613 QNQELOQSLSE-----QLOKENKELRAEAE---LGHELQOAGLTKKEAQTCRH 1661

458 LOTERNDLNKRIV---QDLGAGGQSL--TDSGPERPEGPGCAQAPSSPRVTEAPCPYGA- 511
Db 1662 LTAQVRSLEAQVAHADQQLRDGLKGFQVATDALKSRFPQAK--PQLDLSIDSLDLSCEGTP 1720

512 -----PSTEASQGTGPQEPS 527
Db 1721 LSITSKLPTQPDGTSVPGEPAS 1743

RESULT 31
US-08-466-390-4
Sequence 4, Application US/08466390
Patent No. 5686562
GENERAL INFORMATION:
APPLICANT: TOUTATLY, GARY
APPLICANT: LIDGARD, GRAHAM P
TITLE OF INVENTION: NOVEL MALIGNANT CELL TYPE MARKERS OF THE
NUMBER OF SEQUENCES: 6
CORRESPONDENCE ADDRESS:
ADDRESSEE: TESTA, HURWITZ & THIBEAULT
STREET: 125 HIGH STREET
CITY: BOSTON
STATE: MA
COUNTRY: USA
ZIP: 02110
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: PCT/US93/06160
FILING DATE: 19930621
CLASSIFICATION:
ATTORNEY/AGENT INFORMATION:
NAME: PITCHER ESQ, EDMUND R
REGISTRATION NUMBER: 27,829
REFERENCE/DOCKET NUMBER: MTP-013
TELECOMMUNICATION INFORMATION:

MOLECULE TYPE: protein
US-08-195-487-4

Query Match 8.3%; Score 223; DB 1; Length 2101;
Best Local Similarity 22.1%; Pred. No. 1.2e-06;
Matches 111; Conservative 91; Mismatches 229; Indels 72; Gaps 17;

42 AQARTAGSALRDVSELSQLEEDILSTYCVNNQGGPGBD-GAOGEPAPEDAKSRIV 100
Db 1296 ASGREAEK--QRVASENLRQEL-----TSQAEAEELQELKAWQKFFQKEQAL 1343

101 VARGNEPEPTPVVYGEKPSKGDPTNTEIRQSDVEGDRDHR--POEKKAKGLGKEITL 158
Db 1344 STLQLEHTSTQALVSELLPAK--HLCQQLAQEAQAAAEKRRHELEQSKQAAGLRAELLR 1401

159 LMQTLNLTSTPEEKLAALCKKYAELEHRSNOKMLKQKQOLVQEKDHLRGHESKA 218
Db 1402 AQRELGLIIFERQVAEODRTAQQLRAEKASVAEQLSMLKKAHGLLAEN---RGLGERA 1458

219 VLARSKLESLEL--QRNRSLSKEGVQRAEEREEKKEVTSHPQVTLNDIOLQMSQHNE 277
Db 1459 NLGROFLV---ELDOAREKYVQELAAVRA--DAETRLAEVQREBAQSTARELEVMTAKYEG 1514

278 RNSKLQENMELAEKLLIEQVLEHREHIDKVFHKDLQOQLVDKLOQAQEMLKEABE 337
Db 1515 AKVLEERQRFQERQKLTQVE-----ELSKKLADS-----DQASK 1552

338 RHQREKDFLKEAVESQRMCELMKQETHLKOQALYTEKPEEFQNTLSKSSSEVFTTFKQ 397
Db 1553 VQOQKLKAVQAQGGESQQAQRFQALNELQALSKQEAHVKLQMEKAKTHYDAKKQ 1612

398 EMEKMTKIKKLEKETTMYRSWESSNKALLEWABEKTVRDKELEGLOVKIQLEKLCRA 457
Db 1613 QNQELOQSLSE-----QLOKENKELRAEAE---LGHELQOAGLTKKEAQTCRH 1661

458 LOTERNDLNKRIV---QDLGAGGQSL--TDSGPERPEGPGCAQAPSSPRVTEAPCPYGA- 511
Db 1662 LTAQVRSLEAQVAHADQQLRDGLKGFQVATDALKSRFPQAK--PQLDLSIDSLDLSCEGTP 1720

512 -----PSTEASQGTGPQEPS 527
Db 1721 LSITSKLPTQPDGTSVPGEPAS 1743

RESULT 30
PCT-US93-06160-4
Sequence 4, Application PC/TUS9306160
GENERAL INFORMATION:
APPLICANT:
TITLE OF INVENTION: NOVEL MALIGNANT CELL TYPE MARKERS OF THE
NUMBER OF SEQUENCES: 6
CORRESPONDENCE ADDRESS:
ADDRESSEE: TESTA HURWITZ & THIBEAULT
STREET: 53 STATE STREET
CITY: BOSTON
STATE: MA
COUNTRY: USA
ZIP: 02109
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: PCT/US93/06160
FILING DATE: 19930621
CLASSIFICATION:
ATTORNEY/AGENT INFORMATION:
NAME: PITCHER ESQ, EDMUND R
REGISTRATION NUMBER: 27,829
REFERENCE/DOCKET NUMBER: MTP-013
TELECOMMUNICATION INFORMATION:

CITY: BOSTON
STATE: MA
COUNTRY: USA
ZIP: 02110
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/470,950
FILING DATE: 06-JUN-1995
CLASSIFICATION: 536
ATTORNEY/AGENT INFORMATION:
NAME: PITCHER ESQ, EDMUND R
REGISTRATION NUMBER: 27,829
REFERENCE/DOCKET NUMBER: MTP-013
TELEPHONE: (617) 248-7100
TELEFAX: (617) 248-7000
INFORMATION FOR SEQ ID NO: 4:
SEQUENCE CHARACTERISTICS:
LENGTH: 2101 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-466-390-4

Query Match 8.2%; Score 222; DB 1; Length 2101;
Best Local Similarity 22.1%; Pred. No. 1.4e-06;
Matches 111; Conservative 91; Mismatches 229; Indels 72; Gaps 17;
QY 42 AQAATAGALRDVSELSRQLEDILSTYCVNNGGPGED-GAAGEPAEPDAAKSRTY 100
DB 1296 ASGREAEK--QRVASENLRQEL-----TSQERAEELQELKAWQKFFQKEQAL 1343
QY 101 VARNGEPEPTPVYGEKPSKGDPTNEIRQSEVDRDHR--POEKKAAGLGEITL 158
DB 1344 STLQLEHTSTQALVSELLPAK--HLCQQLQAEQAAAEKRRHELEQSQQAAGLRAELLR 1401
QY 159 LMOTLNTLSTPEEKLAALCKKYAELEHNSOKMKLLQKQSQQLVQEKDHLRGEHSA 218
DB 1402 AQELGELIPLRQKVAEQERTAQQLRAEKASYAEQLSMLKKAHGLIAEEN---RGLGERA 1458
QY 219 VLARSKLESICREL--QRHNSRLKEGVQARAEERKEEVTSHFQVTLNDIQLOMEQHNE 277
DB 1459 NUGROFLEV---ELDQAREKYVQELAAVRA--DAETRLAEVQREAAQSTARBLEVMTAKYEG 1514
QY 278 RNSKLQENMELAEELKXLIQYELREEHIDKVFKHDLQOQLVDAKLQQAQEMLEKAE 337
DB 1515 AKVLEERQRFQERQKLTQAVE-----ELSKKLADS-----DQASK 1552
QY 338 RHOREKDFLLKEAVESQRMCELMKQETHLKOQLALYTEKEFEFQNTLSKSSSEVFTTFKQ 397
DB 1553 VQOQKLKAVQAQGGESQQAQRFQALNELQALSQEQAAYHKLQMEKAKTHYDAKKQ 1612
QY 398 EMERKTKIKKLEKETTMYRSRWESSNKALLEMAEKTVDKLEGLQVKIQLEKLCRA 457
DB 1613 QNQLQEQRLSLE-----QLQENKELAEAR---LGHLEQQAAGLTKAEQTCRH 1661
QY 458 LQTERNDLAKRV----QDLASAGGSL--TDSGPERPPEGCAQAPSSPRVTEAPCPYGA- 511
DB 1662 LTAQVRSLEAQVAHADQQLDLKGFQVATDALKSREPOAK-PQLDLSIDSLSCBEGTP 1720
QY 512 -----PSTEASGQTGPQEPPTS 527
DB 1721 LSITSKLPRTPDGTSGVPEGPAS 1743

RESULT 32
US-08-470-950-4
Sequence 4, Application US/08/470,950
Patent No. 5698439
GENERAL INFORMATION:
APPLICANT: TOUTATLY, GARY
TITLE OF INVENTION: NOVEL MALIGNANT CELL TYPE MARKERS OF THE
TITLE OF INVENTION: INTERIOR NUCLEAR MATRIX
NUMBER OF SEQUENCES: 6
CORRESPONDENCE ADDRESS:
ADDRESS: TESTA, HURWITZ & THIBEAULT
STREET: 125 HIGH STREET

CITY: BOSTON
STATE: MA
COUNTRY: USA
ZIP: 02110
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/470,950
FILING DATE: 06-JUN-1995
CLASSIFICATION: 536
ATTORNEY/AGENT INFORMATION:
NAME: PITCHER ESQ, EDMUND R
REGISTRATION NUMBER: 27,829
REFERENCE/DOCKET NUMBER: MTP-013
TELEPHONE: (617) 248-7100
TELEFAX: (617) 248-7000
INFORMATION FOR SEQ ID NO: 4:
SEQUENCE CHARACTERISTICS:
LENGTH: 2101 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-470-950-4
Query Match 8.2%; Score 222; DB 1; Length 2101;
Best Local Similarity 22.1%; Pred. No. 1.4e-06;
Matches 111; Conservative 91; Mismatches 229; Indels 72; Gaps 17;
QY 42 AQAATAGALRDVSELSRQLEDILSTYCVNNGGPGED-GAAGEPAEPDAAKSRTY 100
DB 1296 ASGREAEK--QRVASENLRQEL-----TSQERAEELQELKAWQKFFQKEQAL 1343
QY 101 VARNGEPEPTPVYGEKPSKGDPTNEIRQSEVDRDHR--POEKKAAGLGEITL 158
DB 1344 STLQLEHTSTQALVSELLPAK--HLCQQLQAEQAAAEKRRHELEQSQQAAGLRAELLR 1401
QY 159 LMOTLNTLSTPEEKLAALCKKYAELEHNSOKMKLLQKQSQQLVQEKDHLRGEHSA 218
DB 1402 AQELGELIPLRQKVAEQERTAQQLRAEKASYAEQLSMLKKAHGLIAEEN---RGLGERA 1458
QY 219 VLARSKLESICREL--QRHNSRLKEGVQARAEERKEEVTSHFQVTLNDIQLOMEQHNE 277
DB 1459 NUGROFLEV---ELDQAREKYVQELAAVRA--DAETRLAEVQREAAQSTARBLEVMTAKYEG 1514
QY 278 RNSKLQENMELAEELKXLIQYELREEHIDKVFKHDLQOQLVDAKLQQAQEMLEKAE 337
DB 1515 AKVLEERQRFQERQKLTQAVE-----ELSKKLADS-----DQASK 1552
QY 338 RHOREKDFLLKEAVESQRMCELMKQETHLKOQLALYTEKEFEFQNTLSKSSSEVFTTFKQ 397
DB 1553 VQOQKLKAVQAQGGESQQAQRFQALNELQALSQEQAAYHKLQMEKAKTHYDAKKQ 1612
QY 398 EMERKTKIKKLEKETTMYRSRWESSNKALLEMAEKTVDKLEGLQVKIQLEKLCRA 457
DB 1613 QNQLQEQRLSLE-----QLQENKELAEAR---LGHLEQQAAGLTKAEQTCRH 1661
QY 458 LQTERNDLAKRV----QDLASAGGSL--TDSGPERPPEGCAQAPSSPRVTEAPCPYGA- 511
DB 1662 LTAQVRSLEAQVAHADQQLDLKGFQVATDALKSREPOAK-PQLDLSIDSLSCBEGTP 1720
QY 512 -----PSTEASGQTGPQEPPTS 527
DB 1721 LSITSKLPRTPDGTSGVPEGPAS 1743
RESULT 33
US-08-467-781-4
Sequence 4, Application US/08/467,781
Patent No. 5780596

GENERAL INFORMATION:
APPLICANT: TOURKATLY, GARY
TITLE OF INVENTION: NOVEL MALIGNANT CELL TYPE MARKERS OF THE
TITLE OF INVENTION: INTERIOR NUCLEAR MATRIX
NUMBER OF SEQUENCES: 6
CORRESPONDENCE ADDRESS:
ADDRESSEE: TESTA, HURWITZ & THIBEAULT
STREET: 125 HIGH STREET
CITY: BOSTON
STATE: MA
COUNTRY: USA
ZIP: 02110

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/467,781
FILING DATE: 06-JUN-1995
CLASSIFICATION: 424

ATTORNEY/AGENT INFORMATION:
NAME: PITCHER ESQ, EDMUND R
REGISTRATION NUMBER: 27,829
REFERENCE/DOCKET NUMBER: MTP-013
TELECOMMUNICATION INFORMATION:
TELEPHONE: (617) 248-7000
TELEFAX: (617) 248-7100

INFORMATION FOR SEQ ID NO: 4:
SEQUENCE CHARACTERISTICS:
LENGTH: 2101 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein

US-08-467-781-4

Query Match 8.2%; Score 222; DB 1; Length 2101;
Best Local Similarity 22.1%; Pred. No. 1.4e-06;
Matches 111; Conservative 91; Mismatches 229; Indels 72; Gaps 17;

QY 42 AQARTAGSALRDVSEELSRQLEDILSTYCVNNQGGPGED-GAQQEPAPEDAEKSRTY 100
DB 1296 ASGREAEK--QRVASENLRLQEL-----TSQAERAEELGQELKAWQKFFQEQAL 1343

QY 101 VARGNEPEPTPVVYGEKPSKGDPTNTEIROSDEVGDRDHR--POEKKKAGLGKEITL 158
DB 1344 STLQLEHTSTQALVSELLPAK--HLCQQLAQEAQAAAEKRRBELEQSKQAAGLRAELLR 1401

QY 159 LMOTLNTLSTPEEKLAALCKKYAEELLEHRNSQKMLQKQKQOLVQEKDHLRGEHSA 218
DB 1402 AQRELGLIPLRQKVABQERTAQQLRAEKASYAEGLSMLKKAHGLLAEN---RGLGERA 1458

QY 219 VLARSKLESICREL--QRHNSRLKEEGVQARAEERBEKKEVTSHFOVTLNDIQLQMEQNE 277
DB 1459 NLGRQFLV---ELDQAREKYVQELAAVRA--DAETRLAEVQREASTARELEVMTAKYEG 1514

QY 278 RNSKLQENNELARLKKLIEQYELREHIDKVFKHDLQOQLVDAKLQQAQEMLKEABE 337
DB 1515 AKVKVLEERQRFQERQKLTAAVE-----ELSKKLADS-----DQASK 1552

QY 338 RHOREKDFLLKEAVESQRMCELMKQETHLKOQLALVTEKFEFNTLSKSSRVFTTQK 397
DB 1553 VQOQKLKAVQAQGGESQQAQRFOAQLNELQAOLSKQEAHYKLQMEKAKTHYDAKKQ 1612

QY 398 EMKMTKKIKKLEKTTMYRSRWESSNKALLEMAEBKTVRDKEGLEQVKIQRLKLCRA 457
DB 1613 QNGELOEQRLSL-----QLQKENKELRAEAER---LGHLEQAGLKTKEAEOCTCRH 1661

QY 458 LQTERNDLKNRV---ODLSAGQGGSL--TDSGPEREPGPGQAQPSPRVTRAPCYPGA- 511
DB 1662 LTAQVRSLEAQAHAQOQLRDGLKGFQVATDLAKSREPQAK--POLDUSIDLSLSCREGTP 1720

QY 512 -----PSTEASGGTGPQBPPTS 527
DB 1721 LSITSKLPRTQPDGTSVPGEPAS 1743

RESULT 34

US-08-483-924-4
Sequence 4, Application US/08483924
Patent No. 5882876
GENERAL INFORMATION:
APPLICANT: TOURKATLY, GARY
TITLE OF INVENTION: NOVEL MALIGNANT CELL TYPE MARKERS OF THE
TITLE OF INVENTION: INTERIOR NUCLEAR MATRIX
NUMBER OF SEQUENCES: 6
CORRESPONDENCE ADDRESS:
ADDRESSEE: TESTA, HURWITZ & THIBEAULT
STREET: 125 HIGH STREET
CITY: BOSTON
STATE: MA
COUNTRY: USA
ZIP: 02110

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/483,924
FILING DATE: 07-JUN-1995
CLASSIFICATION: 435

ATTORNEY/AGENT INFORMATION:
NAME: PITCHER ESQ, EDMUND R
REGISTRATION NUMBER: 27,829
REFERENCE/DOCKET NUMBER: MTP-013
TELECOMMUNICATION INFORMATION:
TELEPHONE: (617) 248-7000
TELEFAX: (617) 248-7100

INFORMATION FOR SEQ ID NO: 4:
SEQUENCE CHARACTERISTICS:
LENGTH: 2101 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein

US-08-483-924-4

Query Match 8.2%; Score 222; DB 2; Length 2101;
Best Local Similarity 22.1%; Pred. No. 1.4e-06;
Matches 111; Conservative 91; Mismatches 229; Indels 72; Gaps 17;

QY 42 AQARTAGSALRDVSEELSRQLEDILSTYCVNNQGGPGED-GAQQEPAPEDAEKSRTY 100
DB 1296 ASGREAEK--QRVASENLRLQEL-----TSQAERAEELGQELKAWQKFFQEQAL 1343

QY 101 VARGNEPEPTPVVYGEKPSKGDPTNTEIROSDEVGDRDHR--POEKKKAGLGKEITL 158
DB 1344 STLQLEHTSTQALVSELLPAK--HLCQQLAQEAQAAAEKRRBELEQSKQAAGLRAELLR 1401

QY 159 LMOTLNTLSTPEEKLAALCKKYAEELLEHRNSQKMLQKQKQOLVQEKDHLRGEHSA 218
DB 1402 AQRELGLIPLRQKVABQERTAQQLRAEKASYAEGLSMLKKAHGLLAEN---RGLGERA 1458

QY 219 VLARSKLESICREL--QRHNSRLKEEGVQARAEERBEKKEVTSHFOVTLNDIQLQMEQNE 277
DB 1459 NLGRQFLV---ELDQAREKYVQELAAVRA--DAETRLAEVQREASTARELEVMTAKYEG 1514

QY 278 RNSKLQENNELARLKKLIEQYELREHIDKVFKHDLQOQLVDAKLQQAQEMLKEABE 337
DB 1515 AKVKVLEERQRFQERQKLTAAVE-----ELSKKLADS-----DQASK 1552

QY 338 RHOREKDFLLKEAVESQRMCELMKQETHLKOQLALVTEKFEFNTLSKSSRVFTTQK 397
DB 1553 VQOQKLKAVQAQGGESQQAQRFOAQLNELQAOLSKQEAHYKLQMEKAKTHYDAKKQ 1612

QY 398 EMERKTKKIKLEKETMYRSRNSSNKALLEAEKTVDRKLEGLQVKIQRLEKLCRA 457
Db 1613 QNOELOQLRSLE-----QLQKENKELRAEAER---LGHELQAGLTKKEAQTCRH 1661
QY 458 LQTERNDLNKRV-----QDLGAGGGSL-TDGGPERRPBGCAQAPSSPVTRPACVPGA- 511
Db 1662 LTAQVRSLAQVAHADQDLGKGFQVATDALKSRPQAK-PQLDLSIDSLDSCBEGTP 1720
QY 512 -----PSTASGQTGPBPPTS 527
Db 1721 LSITSKLPRTPDQTSVPGEPAS 1743
RESULT 35
US-09-452-294-1
; Sequence 1, Application US/09452294
; Patent No. 6287790
; GENERAL INFORMATION:
; APPLICANT: Lelievre, Sophie
; APPLICANT: Bissell, Mina
; TITLE OF INVENTION: UTILIZATION OF NUCLEAR STRUCTURAL PROTEINS FOR TARGETED
; TITLE OF INVENTION: THRAPAT AND DETECTION OF PROLIFERATIVE AND
; TITLE OF INVENTION: DIFFERENTIATION DISORDERS
; FILE REFERENCES: IB-1454- Sequence Submittal
; Patent No. 6287790
; CURRENT APPLICATION NUMBER: US/09/452,294
; CURRENT FILING DATE: 1999-11-30
; PRIOR APPLICATION NUMBER: 60/110,420
; PRIOR FILING DATE: 1998-11-30
; NUMBER OF SEQ ID NOS: 1
; SOFTWARE: Patent In Ver. 2.1
; SEQ ID NO 1
; LENGTH: 2101
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-452-294-1
Query Match 8.2%; Score 222; DB 3; Length 2101;
Best Local Similarity 22.1%; Pred. No. 1.4e-06;
Matches 111; Conservative 91; Mismatches 229; Indels 72; Gaps 17;
QY 42 AQARTAGSALRDVSRLESHQEDILSTYCVNDNQGPGED-GAQGEPAEPEDAEKSRTY 100
Db 1296 ASGREAEK--QRVASENLRLQEL-----TSQARAEFLGQELKAWQEPFOKEQAL 1343
QY 101 VARNGPEPTFVYGEKPSKGPNTTEIRQSDVGDROHRR--PQKKKAKGLGKEITL 158
Db 1344 STQLQHTSTQALVSELLPAK--HLCOQLAQEAQAAEKHRELEQSKQAAGGLRAELIR 1401
QY 159 LMQTLNTLSTPEKLAALCKKYAELEHNSQKMKLLQKQSQVLQVQEKHRLGHEHSA 218
Db 1402 AQRELGLPIRLQKVAEQERTAQQLRAEKASVAEQSLMKKAGHLLAEEN---RGLGERA 1458
QY 219 VLARSKLESICREL-QRHNRSLEKEGVQARAEERKEKVTSHFVTLNDIQLQXEQHNE 277
Db 1459 NLGRQFLEV---ELDQAREKYVQELAAVRA-DRETRLAEVQEAQSTARELVMTAKTEG 1514
QY 278 RNSKLQENNELAERLKUKLIEQVELREEHIDKVFQKHDLQOOLVDAKLOQAQEMLEKAE 337
Db 1515 AKVVKLEERQRFQERQKLTQAVE-----ELSKKLADS-----DQASK 1552
QY 338 RHQREKDFLLKAVESORNCMLKQSTHLKQOALALYTEKPEFPQNTLSKSSEVFTFKQ 397
Db 1553 VQOQKLKAVQAQGESQQAQRFQALNELQALQSKQEAABHYKLQMSKATHYDAKQ 1612
QY 398 EMERKTKKIKLEKETMYRSRNSSNKALLEAEKTVDRKLEGLQVKIQRLEKLCRA 457
Db 1613 QNOELOQLRSLE-----QLQKENKELRAEAER---LGHELQAGLTKKEAQTCRH 1661
QY 458 LQTERNDLNKRV-----QDLGAGGGSL-TDGGPERRPBGCAQAPSSPVTRPACVPGA- 511
Db 1662 LTAQVRSLAQVAHADQDLGKGFQVATDALKSRPQAK-PQLDLSIDSLDSCBEGTP 1720

QY 512 -----PSTASGQTGPBPPTS 527
Db 1721 LSITSKLPRTPDQTSVPGEPAS 1743
RESULT 36
US-08-328-254-6
; Sequence 6, Application US/08328254
; Patent No. 5710022
; GENERAL INFORMATION:
; APPLICANT: Zhu, Xueliang
; APPLICANT: Lee, Wen-Hwa
; TITLE OF INVENTION: A No. 5710022el Nuclear Mitotic Phosphoprotein
; NUMBER OF SEQUENCES: 8
; CORRESPONDENCE ADDRESS:
; ADDRESSES: Campbell and Flores
; STREET: 4370 La Jolla Village Drive, Suite 700
; CITY: San Diego
; STATE: California
; COUNTRY: USA
; ZIP: 92122
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION NUMBER: US/08/328,254
; FILING DATE: 24-OCT-1994
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/141,239
; FILING DATE: 22-OCT-1993
; ATTORNEY/AGENT INFORMATION:
; NAME: Campbell, Cathryn A.
; REGISTRATION NUMBER: 31,815
; REFERENCE/DOCKET NUMBER: P-CJ 1191
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (619) 535-9001
; TELEFAX: (619) 535-8949
; INFORMATION FOR SEQ ID NO: 6:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2482 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-328-254-6
Query Match 8.2%; Score 220.5; DB 1; Length 2482;
Best Local Similarity 24.2%; Pred. No. 2.1e-06;
Matches 127; Conservative 79; Mismatches 192; Indels 126; Gaps 24;
QY 106 RPEPTFVYGEKPSKGPNTTEIRQSDVGDROHRRPQEKKAK--GLGKEITLL----- 159
Db 1792 EKEQKVQMKESSTAMEMLOTQLKELNVAALH-NDQBACKAQNLSSQVCELEBK 1850
QY 160 MQTLNTLSTPEKLAALCKKYAELEHNSQKMKLLQKQ-----QSLVQ 206
Db 1851 AQLLQGLDEAKNNYIVLQSSVNGLIQVEDGQK--LEKKDETSRLKNQIQDQBLVS 1907
QY 207 EKHRLGHEH---SKAVLARS---KLSLCELQENHRSLEK--EGVQARAEERKEKVEV 257
Db 1908 KLSQVEGEHQLWKQCNDLRLNLTVELSOKIQLVQSKNASLQDTLEVLQSSYKNELEL 1967
QY 258 TSH-----FOVTLNDIQLQXEQHNE-----ERNSKLQENNELAERLKLCLEQY-- 300
Db 1968 TKMDKMSPVKVNKMTAKETELQREHMEHAQTAELQELSGEKNRLAGELQLLEBK 2027
QY 301 -----ELREE-----HIDKVPKHDLQOOLVDAKLOQAQEMLEKAEERHQ---- 340
Db 2028 SKQQLKEITLSELSKSLDCHQKQVEKEGKVEETAEYQLR-----LHEAKGHQALL 2082

QY 341 -----REKDFLLKEAVESORM-CELMKQOETHLKQOLALYTEKTEFBFQNTL 395
DB 2083 LDTNKQVEIOTYREKLTKEECLSSQKLEIDLKSSKEELNNSLKATTOHLEELKTK 2142
QY 386 SKSESVPTTFKQEMKEMKTKIKLEKETTWYRSRWSSNKAL--LEMAEK-----TVRDK 439
DB 2143 MDNLKYVNLKKNERRAQGRMKLLKSKCKQLEBEKEITLOKELSOQAQEKQKTGTVMDT 2202
QY 440 ELSEGLQVRIOR-----LEKLC-----RALOTERDNLKRVQDLSAGGQ 478
DB 2203 KVELDTTEIKELKETLEETKEADEYLDKYCSLLISHKLEKAKEMLETOVAHLCS--QQ 2260
QY 479 SLTDS--GPRRRPEGCAQAPSS--PRVTEAPCYPGAPSTEASGO 519
DB 2261 SKQDSRGSLPGWPFG---PSPISPVTEKELSSG--QNKASCK 2299

RESULT 37
US-09-866-108A-15753
; Sequence 15753, Application US/09866108A
; Patent No. 6686188
; GENERAL INFORMATION:
; APPLICANT: GU, Yizhong
; APPLICANT: JI, Yonggang
; APPLICANT: PENN, Sharon G.
; APPLICANT: HANZEL, David K.
; APPLICANT: RANK, David R.
; APPLICANT: CHEN, Wensheng
; APPLICANT: SHANNON, Mark
; TITLE OF INVENTION: MYOSIN-LIKE GENE EXPRESSED IN HUMAN HEART AND MUSCLE
; FILE REFERENCE: AROMICA-7
; CURRENT APPLICATION NUMBER: US/09/866,108A
; CURRENT FILING DATE: 2001-05-25
; PRIOR APPLICATION NUMBER: US 60/207,456
; PRIOR FILING DATE: 2000-05-26
; PRIOR FILING DATE: 2000-10-04
; PRIOR APPLICATION NUMBER: US 60/236,359
; PRIOR FILING DATE: 2000-09-27
; PRIOR APPLICATION NUMBER: PCT/US01/00666
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00667
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00664
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00669
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00665
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00668
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00663
; PRIOR FILING DATE: 2001-01-30
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 15755
; SOFTWARE: Aromica Sequence Listing Engine
; Patent No. 6686188
; SEQ ID NO 15753
; LENGTH: 1695
; TYPE: PRT
; ORGANISM: Mus musculus
US-09-866-108A-15753

Query Match 8.1%; Score 219.5; DB 4; Length 1695;
Best Local Similarity 21.8%; Pred No. 1.5e-06;
Matches 123; Conservative 86; Mismatches 219; Indels 137; Gaps 20;
QY 19 SQAAPAVEAPGSSQAPRPEGAQARTAGSALRDVSELSQLEBILSTYCVDMNQG 78
DB 1169 TQKWSLEALBQISSQESKDEASLAKVKQ--LRDLEAKVKQESBEL-----DEQAG 1219
QY 79 PGEDGAGEPABEDAEKSR--TYVARNGBPEPTVYVYGEKPSKGPNTTEIRQ----- 131

DB 1220 SIOMLEQAKLRLEMEMERNQTH-----SKEMESRDEVEEAEARQSCQK 1264
QY 132 -----SDEVGRD-----HRRPQEKKAKGLGKEI 156
DB 1265 KQMEVQLEBEYEDKQALREKRELESKLSLSDQVNRQDFSEKELRDLKTKALLADA 1324
QY 157 TLLMTQNTLTIPTPEKLAALCKYAEELLEHHNSQKMKLQKQKQSQLVQEKDHLRGHS 216
DB 1325 QIMDLKNNAPSKKEIAQLKQ-----LEE-----SEFTCAAAVAKRAMEVEEDLHLD 1377
QY 217 KAVLARSKLSICRELQHRNRSKLEEGVQRAPEEERKEV-----TSHFQVTLN 266
DB 1378 DIAKAKTALBEEQLSRLQRE-----KNEIQNRLEEDQDMXELMKKHKAAVAQASRDMAQN 1433
QY 267 DILOMEQHNEENSKLQENMELAKLLEIQLYELREHIDKVFKEKDLQOOLQADAKLO 326
DB 1434 DLQAQIERSN-----KEKQELQKQALQSQVEF-----LEQSMWDKSLV 1473
QY 327 QAQE--MLXAEABERHOREKDFLLKAEVQSORMCMLKQOETHLKQOLALYTEKTEFBFQNTL 385
DB 1474 SRQEKIRLETRLEFTEKTQVKRLENLASRLKETWEKLTBERDQAA--AENREKEQN-- 1529
QY 386 SKSSEVPTTFKQEMKEMKTKIKLEKETTWYRSRWSSNKALLEMAEKTVPDKELEGLO 445
DB 1530 KQLQRLDRTKEESELARK-----EABASRKKGHELEMDLES--LEAANGSLQAD-----LK 1579
QY 446 VKIORLEKLCRALOTER-----NDLNKRVQDLSAGQGGSLT--DSGPERPE 490
DB 1580 LAFKIGDLOAIBDEMSESDENEDLINSEGDSDVSELEDVGVKSWLSKNKGFSKAPS 1639
QY 491 GPGAQAPSSPRVTEAPCYPGAPSTE 515
DB 1640 DDGSLKSSSPTSHWKPLAPDPDSDE 1664

RESULT 38
US-08-795-475-6
; Sequence 6, Application US/08795475
; Patent No. 5965390
; GENERAL INFORMATION:
; APPLICANT: Bivrick, Lars
; APPLICANT: Sivbring, Ulf
; TITLE OF INVENTION: PROTEIN L AND HYBRID PROTEINS THEREOF
; NUMBER OF SEQUENCES: 14
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SEED and BERRY LLP
; STREET: 6300 Columbia Center, 701 Fifth Avenue
; CITY: Seattle
; STATE: Washington
; COUNTRY: USA
; ZIP: 98104-7092
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/795,475
; FILING DATE: 11-FEB-1997
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Mcmasters, David D.
; REGISTRATION NUMBER: 33,963
; REFERENCE/DOCKET NUMBER: 100084.402D1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (206) 622-4900
; TELEFAX: (206) 682-6031
; INFORMATION FOR SEQ ID NO: 6:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 443 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein

```
US-08-795-475-6
Query Match      8.1%; Score 218; DB 2; Length 443;
Best Local Similarity 23.4%; Pred. No. 3.6e-07;
Matches 111; Conservative 86; Mismatches 178; Indels 100; Gaps 21;

QY 104 NCEPPTVVYGEKSKDPNTEIR-----QDEVGDRDHRPQKKKAK 150
DB 1 NGDGNPREVI---EDUAAANPAIOTRLRHEHVKDLARLENAMVAGRDFKAEHLEKAK 57
QY 151 GLGKEITLQMOTL-NTLSTPEEKLAALCKKYAELEHNSOKOM--KLQKKQSQLOVE 207
DB 58 -----QALEQDKDLETGKELQDY-DLAKESTWDRQRLKELEBEKKEALEA 106
QY 208 KQHLRGEHSAVLAARSKLSLREL-----ORHNR-SLKEEGVQARAEKEKEKVT 258
DB 107 IQASDYHRAALEKELEKKALELAIDQASQDYNRANVLEKELETTITRQEIINRL 166
QY 259 SHFQVTLNDIOLQMEQHNRNSKLRQ-----NMELAERLKLIQ--YELR 303
DB 167 GNAKLELDQSSKEQTLTEKAKLEEKQISDASRQSLRDLDSAREAKKQVEKOLANT 226
QY 304 EKHIDKVFHKDLOQLQVDAKQQAQEMLEKABE-RHQBKQFLLEAVESSQKMLKQ 362
DB 227 AE-LDKVKEDK-----QISDASRQSLRDLDSAREAKKQVEKOL- 273
QY 363 QETHLQQLALYTEKEFEFQNTLSKSEVFTTFKQEMKMTKKIKKLEKETTMYRSMES 422
DB 274 -----VKEKQISDASRQSLRDLDSAREAKKQVEKALEEANSKLALEK----- 318
QY 423 SNKALLEMAEKTVDKLEGLQVKIQRLEKLCRALQTERNDLNKQVQDSAGGQSLTD 482
DB 319 LNK-----ELSEKSKLTKEXAELOAK--LEASAKAL---KEOLAKQAEIARLAKGKASD 369
QY 483 S-GPERRP-----EGPQAQAPSSPRVTEAPCYPGAPSTEASGTPQEPPTSAR 529
DB 370 SQTPDTPKGNKAVPGKQAPQAGTKPNQKAPMKETKQPLPSTGETANPFFTAAR 424

RESULT 39
US-09-914-259-21
; Sequence 21, Application US/09914259
; Patent No. 6495336
; GENERAL INFORMATION:
; APPLICANT: Makowski, Lee
; APPLICANT: Hyman, Paul
; APPLICANT: Williams, Mark
; TITLE OF INVENTION: STAGED ASSEMBLY OF NANOSTRUCTURES
; FILE REFERENCE: 8471-010-999
; CURRENT APPLICATION NUMBER: US/09/914,259
; CURRENT FILING DATE: 2000-11-21
; NUMBER OF SEQ ID NOS: 180
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 21
; LENGTH: 967
; TYPE: PRT
; ORGANISM: Loligo pealeii
US-09-914-259-21

Query Match      8.0%; Score 217.5; DB 4; Length 967;
Best Local Similarity 20.0%; Pred. No. 1e-06;
Matches 110; Conservative 95; Mismatches 245; Indels 101; Gaps 17;

QY 48 QSGALRDVSELSRQLEDIL--STYCDVNNQGGPGEDGAGPEADEKSRITYAENG 105
DB 436 QSOLIEKLEQOMEQ-EDLIQARRRYENILQDMSRIQADNSAKDEVKEVQAL----- 489
QY 106 EPSPFTPVYGEKPSKDPNTEIROSDEVGDR-----DHR- 142
DB 490 --BELAMVDQKQVEDKDNKENENLSEELNQKLSLNLQNELDQLKQSSXHKRVRTD 547
QY 143 -----POEKKKAKGLGKE-----ITLLMQTLNLTLSPEK 172
```

146 -----KKKAGLGKEITLLMTLSTP-----EBKLAALCKYAEELLEHRN 189
Db ARQKLEKVTAAKIKLEDDILVMDQNNKSKERKLEERISLT---TNLAEEBK 1021
190 SQKMKLQKQSQVOEKDHLRGEHSAVLARSKLESICRLQRNRSLEK--- 242
Db AKNLTAKNKHESMISELVRLAKEE---SRQLEKLRKXMDGEASDLHEQIADLQAO 1077
243 ---GVQAREEBE-----KXKVTSHFQVTLNDIQLQMEQHNER 278
Db IAEKMLQAKSELEQAALAEDETSQKXNALKIRLEGH-----ISLDQEDLSERAA 1133
279 NSKLQENNELAERLKKLIEQVELREHIDKVFHKDLOOQLVDAKLAQAQENLKEABER 338
Db RNKAEQKRDGLGLEAL---KTEL---EDTLTTATQOELR-----AKREQEVTLKKALDE 1185
339 HQEKDFLLKAEVESQRMCELMKQOETHLKQALALATKFEER-----ONTLSK 388
Db ETRASHAQOVB-----MQKQTVVEEL---TEQLEQPKRAKAMLDKTKOTLEKE 1232
389 S-----EYFTTFKQEMERTYKIKKLEKETTMYESRWESSNKA----- 426
Db NADLAGELRVLQAKQVE---HKKKLEVLQELQSLQKSDGERARAEMLNDKVHKLQNEV 1289
427 -----LLEMAEKTVR-DKELEGLOVKIOLEKL-----CRALQTERNDLN 466
Db BSVTGMLSEAGRAIKLAKEVASLGSQQLQDTQSELQEBTRQKLVNSTKURQLEDERNSLQ 1349
467 KRYQD 471
Db EQLDE 1354

RESULT 41

US-08-533-306A-6
; Sequence 6, Application US/08533306A
; Patent No. 5837457
; GENERAL INFORMATION:
; APPLICANT: Liu, Pu
; APPLICANT: Collins, Francis S.
; APPLICANT: Siciliano, Michael J.
; APPLICANT: Claxton, David
; TITLE OF INVENTION: Markers for Detection of Chromosome 16
; TITLE OF INVENTION: Rearrangements
; NUMBER OF SEQUENCES: 14
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Harness, Dickey & Pierce, P.L.C.
; STREET: P.O. Box 828
; CITY: Bloomfield Hills
; STATE: MI
; COUNTRY: USA
; ZIP: 48303
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/533.306A
; FILING DATE: September 25, 1995
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Smith, DeAnn F.
; REGISTRATION NUMBER: 36683
; REFERENCE/DOCKET NUMBER: 2115-00869COB
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (810) 641-1600
; TELEFAX: (810) 641-0270
; INFORMATION FOR SEQ ID NO: 6:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 816 amino acids
; TYPE: amino acid
; TOPOLOGY: linear

MOLECULE TYPE: protein
US-08-533-306A-6

Query Match 8.0%; Score 215.5; DB 2; Length 816;

Best Local Similarity 22.8%; Pred. No. 1.2e-06;
Matches 129; Conservative 80; Mismatches 218; Indels 139; Gaps 20;

QY 26 EAGPG-----SSQAPRKPEGAQARTAQ-----SGALRDVSE-----E 58
Db 176 EAGKRAIKLANDVASLSQQLDTQELLQEBTRQKLVNSTKLRQLEEBENSLODQLDEME 235
QY 59 LSRQLEDILSYCYVNNQCGPGEDGAQGEPAEPDAEKSRITYVARNGSP-----BPTPVV 113
Db 236 AKONLERHISTL---NIOLSKKKLQDPASTVBALEEGKRGKRIENLTCQYHEKAAA 292
QY 114 YGEKPSKGDPTBEIRQSDVGRDHRHPQEKKAKGLKEITLLMTLSTL----- 167
Db 293 YDKLETK---NRLQQLDLDLVLDNQRQVSNLEKKQRKFDQLLAEEKNISSKYADER 349
QY 168 -----TPBEKLAALCKYAEELLEHRNSOKOMKLLQKQSQVQEKHRLGHEHSAV 219
Db 350 DRAEABAREKETKALSLARALEEALAEKEELERTNMKLAEMEDLVSSKDDVGRNVHLE 409
QY 220 LARSKLBSLCHLOHNRSLKEBQVQARAREEBEERKKEVT-----SHFQVTLNDIQLQMEQ 274
Db 410 KSKRALETQMEEMTKQLELEDE---LQASEDAKLRLVNMQALKGQFB---RDLQARDQ 464
QY 275 HNRNSKURQENME-----LAERLKKLIE---QYEL-----REEHIDK 309
Db 465 NEEKERQLOQLHVEYTELEDERNERALAAAKKLEGLDKLEQLQADSATKGRBEALNQ 524
QY 310 VFK---HKDLQOQLVDAK-----LQQAQEMLKEABERHOR- 341
Db 525 LKQQAQKMDFORELEDAERDEIPATAKENEKAKSLEADLMQLQEDLAAAEARQA 584
QY 342 ---EKDFLLKAEVESQRMCELMKQOETHLKQALALATKFEERFQNTLSKSESVFTTFKQEM 399
Db 585 DIEKEBELAEELASSLGERNALQDEKRLAEARTAQLEEBEEEEQGNMEASD-----RV 637
QY 400 EKKTKKIKKLEKETTMYESRWESSNKALLEM-ABEKTVRDK---ELEG-----L 444
Db 638 RQATQOASQLSNELATERSTQKNESARQQLERQNKELRKLHEMEGAVSKFKSTIAAL 697
QY 445 QVKIQRLKLCALQTERNDLNKRVQ 470
Db 698 EAKIAQLEBQVEQBARERQAATKSLK 723

RESULT 42

US-08-742-923A-6
; Sequence 6, Application US/08742923A
; Patent No. 5869611
; GENERAL INFORMATION:
; APPLICANT: Liu, Pu
; APPLICANT: Collins, Francis S.
; APPLICANT: Siciliano, Michael J.
; APPLICANT: Claxton, David
; TITLE OF INVENTION: Markers for Detection of Chromosome 16
; TITLE OF INVENTION: Rearrangements
; NUMBER OF SEQUENCES: 14
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Harness, Dickey & Pierce, P.L.C.
; STREET: P.O. Box 828
; CITY: Bloomfield Hills
; STATE: MI
; COUNTRY: USA
; ZIP: 48303
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:

```

; APPLICATION NUMBER: US/08/742,923A
; FILING DATE: NO. 5869611ember 1, 1996
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Smith, DeAnn F.
; REGISTRATION NUMBER: 36683
; REFERENCE/DOCKET NUMBER: 2115-00869DVC
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (810) 641-1600
; TELEFAX: (810) 641-0270
; INFORMATION FOR SEQ ID NO: 6:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 816 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-742-923A-6

Query Match      8.0%; Score 215.5; DB 2; Length 816;
Best Local Similarity 22.8%; Pred. No. 1.2e-06;
Matches 129; Conservative 80; Mismatches 218; Indels 139; Gaps 20;

QY 26 EAEQPG-----SSQAPRKPEGQAQTAQ-----SGALRDVSE-----E 58
Db 176 EAEQKAIKLVASLSQLODTQELLQETROKLVNSTKRLQLEERNLSQDQLEEME 235
QY 59 LSRQLEDILSTYCVNNQGGEDGAGQEPAPEDAKSRVTYVANGEP-----BPTPV 113
Db 236 AKONLERHISTL---NIQSDSKKKLQDFASTVLEAEKGKRFQKEIENLTQYBEKAAA 292
QY 114 YGEKPSKGDPTNTEIRQSDVEGDRHRRPOEKKAKGLGKEITLLMOTLNTLS----- 167
Db 293 YDKLEKTK---NRLQQLDLDLVLDNQRLQVSNLEKKQKQKFDQLLAEBKNTSSKYADER 349
QY 168 -----TPBEKLAALCKKYAEILLLEHNSQKQKLLQKQSOQLVQEKDHLRGHSHKAV 219
Db 350 DRAEABAREKETKALSLARALEEALAEKEELERTNMLKAEMEDLVSSKDDVGNVHLE 409
QY 220 LARSKLESICRELQHNRSKKEEGVQARAREEERKEVT-----SHFQVTLNDIQLOMEQ 274
Db 410 KSKALETQMEBMTQLELEDE--LQASEDAKLRLVNMQALKQGF-----RDLQARDEQ 464
QY 275 HNRNSKLQENME-----LAERLKLIE-----QVEL-----REHIDK 309
Db 465 NEEKRQQLRQLHEYTELEDERNERALAAAKKGLGDLKDLQADSAIKGREALIKQ 524
QY 310 VFK-----HKDLQQLVDAK-----LQQAQEMLKEAEERHOR- 341
Db 525 LRKLAQMKDFQRELEDAASRDEIFATAKENEKKAKSLEADLMQLQEDLAAERAKQA 584
QY 342 --EKDFLLKEAVESQRCMLKQOETHLKKQALALYTEKFEFFQNTLSKSSEVFTTFKQBM 399
Db 585 DLEKEELABELASSLSGRNALQDEKRLRLAIAQLAELEEELEEQGNMEAMSD-----RV 637
QY 400 EYMTKKIKKLEKTTMYRSWESSNKALLEM-AEETVTRDK--ELEG-----L 444
Db 638 RKATQQAQQLSNELATERSTAQKNESARQQLERQNKELSKLHEMEGAVKSKFKSTIAAL 697
QY 445 QVKIQRLEKLCRALQTERNDLNRVQ 470
Db 698 EAKIAQLEQVEQAEKQAATKSLK 723

```

RESULT 43

US-08-533-306A-4

; Sequence 4, Application US/08533306A

; Patent No. 5837457

; GENERAL INFORMATION:

; APPLICANT: Liu, Pu

; APPLICANT: Collins, Francis S.

; APPLICANT: Siciliano, Michael J.

; APPLICANT: Claxton, David

; TITLE OF INVENTION: Markers for Detection of Chromosome 16

```

; TITLE OF INVENTION: Rearrangements
; NUMBER OF SEQUENCES: 14
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Hazness, Dickey & Pierce, P.L.C.
; STREET: P.O. Box 828
; CITY: Bloomfield Hills
; STATE: MI
; COUNTRY: USA
; ZIP: 48303
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/533,306A
; FILING DATE: September 25, 1995
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Smith, DeAnn F.
; REGISTRATION NUMBER: 36683
; REFERENCE/DOCKET NUMBER: 2115-00869COB
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (810) 641-1600
; TELEFAX: (810) 641-0270
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 885 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-533-306A-4

```

```

Query Match      8.0%; Score 215.5; DB 2; Length 885;
Best Local Similarity 22.8%; Pred. No. 1.3e-06;
Matches 129; Conservative 80; Mismatches 218; Indels 139; Gaps 20;

QY 26 EAEQPG-----SSQAPRKPEGQAQTAQ-----SGALRDVSE-----E 58
Db 245 EAEQKAIKLVASLSQLODTQELLQETROKLVNSTKRLQLEERNLSQDQLEEME 304
QY 59 LSRQLEDILSTYCVNNQGGEDGAGQEPAPEDAKSRVTYVANGEP-----BPTPV 113
Db 305 AKONLERHISTL---NIQSDSKKKLQDFASTVLEAEKGKRFQKEIENLTQYBEKAAA 361
QY 114 YGEKPSKGDPTNTEIRQSDVEGDRHRRPOEKKAKGLGKEITLLMOTLNTLS----- 167
Db 362 YDKLEKTK---NRLQQLDLDLVLDNQRLQVSNLEKKQKQKFDQLLAEBKNTSSKYADER 418
QY 168 -----TPBEKLAALCKKYAEILLLEHNSQKQKLLQKQSOQLVQEKDHLRGHSHKAV 219
Db 419 DRAEABAREKETKALSLARALEEALAEKEELERTNMLKAEMEDLVSSKDDVGNVHLE 478
QY 220 LARSKLESICRELQHNRSKKEEGVQARAREEERKEVT-----SHFQVTLNDIQLOMEQ 274
Db 479 KSKALETQMEBMTQLELEDE--LQASEDAKLRLVNMQALKQGF-----RDLQARDEQ 533
QY 275 HNRNSKLQENME-----LAERLKLIE-----QVEL-----REHIDK 309
Db 534 NEEKRQQLRQLHEYTELEDERNERALAAAKKGLGDLKDLQADSAIKGREALIKQ 593
QY 310 VFK-----HKDLQQLVDAK-----LQQAQEMLKEAEERHOR- 341
Db 594 LRKLAQMKDFQRELEDAASRDEIFATAKENEKKAKSLEADLMQLQEDLAAERAKQA 653
QY 342 --EKDFLLKEAVESQRCMLKQOETHLKKQALALYTEKFEFFQNTLSKSSEVFTTFKQBM 399
Db 654 DLEKEELABELASSLSGRNALQDEKRLRLAIAQLAELEEELEEQGNMEAMSD-----RV 706
QY 400 EYMTKKIKKLEKTTMYRSWESSNKALLEM-AEETVTRDK--ELEG-----L 444
Db 707 RKATQQAQQLSNELATERSTAQKNESARQQLERQNKELSKLHEMEGAVKSKFKSTIAAL 766

```


[illegible]

RESULT 45
US-08-685-871-2
? Sequence 2, Application US/08685871
? Patent No. 6013499
? GENERAL INFORMATION:
? APPLICANT: NARUMIYA, Shuh
? APPLICANT: IWAMATSU, Akihito
? TITLE OF INVENTION: RHO TARGET PROTEIN KINASE P160
? NUMBER OF SEQUENCES: 68
? CORRESPONDENCE ADDRESS:
? ADDRESSEE: Foley & Lardner
? STREET: 3000 K Street, N.W., Suite 500
? CITY: Washington
? STATE: D.C.
? COUNTRY: USA
? ZIP: 20007-5109
? COMPUTER READABLE FORM:
? MEDIUM TYPE: Floppy disk
? COMPUTER: IBM PC compatible
? OPERATING SYSTEM: PC-DOS/MS-DOS
? SOFTWARE: Patent In Release #1.0, Version #1.30
? CURRENT APPLICATION DATA:
? APPLICATION NUMBER: US/08/685,871
? FILING DATE: 24-JUL-1996
? CLASSIFICATION: 435
? PRIOR APPLICATION DATA:
? APPLICATION NUMBER: JP 8-184102
? FILING DATE: 25-JUN-1996

PRIOR APPLICATION DATA: JP 7-262553
 APPLICATION NUMBER: 14-SEP-1995
 FILING DATE: 14-SEP-1995
 ATTORNEY/AGENT INFORMATION:
 NAME: BENT, Stephen A.
 REGISTRATION NUMBER: 29,768
 REFERENCE/DOCKET NUMBER: 16887/845
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: (202)672-5300
 TELEFAX: (202)672-5399
 TELEX: 904136
 INFORMATION FOR SEQ ID NO: 2:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 1354 amino acids
 TYPE: amino acid
 TOPOLOGY: linear
 MOLECULE TYPE: protein
 US-08-6395-871-2

Query Match 8.0%; Score 215.5; DB 3; Length 1354;
 Best Local Similarity 25.2%; Pred. No. 2.2e-06;
 Matches 101; Conservative 64; Mismatches 145; Indels 91; Gaps 16;
 127 BEIRQSDVGDHRHPQKKKAKGIGKEITLLIMQTLNTLSTPEEKLAALCKKYAELEE 186

Db 487 EKKMLQHRINEYQKAEQENKERNVNEVSTLKQLEDL-----KKVSQ---- 531

Qy 187 HRNSQKMKLQKQSQOLVQEKHGEHSHKAVLAR-----SKLESCEBLOHRN 236

Db 532 --NSQLANEKLSQKQLEBEANDLLRTESDFAVLKRSHTKSTEMSKSISQLESARELOHRN 589

Qy 237 RSLKEGVQVARE-----EHEKR-----KEVTSHFQVTLNDIQLOME--OHN-----E 277

Db 590 RILENSKSTQDXDYVQALAEARRDRCHDSEMGIDQARITSLQBEVGHKXHNLEKVE 649

Qy 278 RNSKLQENWELAEARKLIEQVELREHIDKVPKHDLOQOL-----VDKLOQA 328

Db 650 GERKADMLNHSKEKXNNLE-----IDLNYKLSLQOORLEQVEVNEHKVTKARLTDK 701

Qy 329 QEMLEKEERHOREKDFLLKEAVESQRMCELMKQOETHLKQALALYTEKPEEPQNTLSKS 388

Db 702 HQSIEBAKSVAMCEMEKLEAREAREKAE---NRVVQIEKQCSMLDVLKQSQKL--- 755

Qy 389 SEVFTTQKQEMKTKKIK-KLEKBTMYRSRSHSSNKALLEMAHEHT--VRQKELEGLO 445

Db 756 -EHLTGKRMEDVKNLTQLQEOE-----SNKRLQLQNELKTAQFEADNLKGLE 804

Qy 446 ----VKIQRLEKLCRALQTERNDLNKRVQDLSAGGGSGLTD 482

Db 805 KQKQENTLLEAKGLLEFFLAQLTKQYR-----GNEQVRE 841

RESULT 46

US-09-976-594-296

; Sequence 296, Application US/09976594

; Patent No. 6673549

; GENERAL INFORMATION:

; APPLICANT: Furness, Michael

; APPLICANT: Buchbinder, Jenny

; TITLE OF INVENTION: GENES EXPRESSED IN C3A LIVER CELL CULTURES TREATED WITH STEROIDS

; FILE REFERENCE: PA-0041 US

; CURRENT APPLICATION NUMBER: US/09/976,594

; CURRENT FILING DATE: 2001-10-12

; PRIOR APPLICATION NUMBER: 60/240,409

; PRIOR FILING DATE: 2000-10-12

; NUMBER OF SEQ ID NOS: 1143

; SOFTWARE: PERL Program

; SEQ ID NO 296

; LENGTH: 1388

; TYPE: PRT

; ORGANISM: Homo sapiens

; FEATURE:

; NAME/KEY: misc_feature

; OTHER INFORMATION: Incyte ID No. 6673549 064987CD1

US-09-976-594-296

Query Match 8.0%; Score 215; DB 4; Length 1388;

Best Local Similarity 21.3%; Pred. No. 2.4e-06;

Matches 132; Conservative 84; Mismatches 173; Indels 230; Gaps 23;

Qy 91 PEDAERS-----TYVARNGEPE-----PTPVV--- 113

Db 316 PEDAISKHAKNLICAFITDREVLGRNGVEIRQHPPKNDQWHDNIRETAAFPVPEL 375

Qy 114 -----YGEKESKGDNPTEEI-----RQDEVG 136

Db 376 SSDIDSNFDDIEDDKGDEVETFPKPAFVGNQLPFIQTYRENLLSDSPSCRENDISQ 435

Qy 137 DRDHRRPQE-----KKKAGLKGKBITLLMQTL 163

Db 436 SRKNEESQEIQKLYTLLEHLSNEMQAKELBQCKSVNTLEKTAKELEEEITLRKSV 495

Qy 164 NTLSTPEKLAALCKKVAEL-----LEEHRNSOK-QMKLLQVK- 200

Db 496 SALRQLREKALLQKNAVQKADHAKRNLNDVNSLKQOLEDLKKNQNSQISTE 555

Qy 201 -----QSOLVQEKDHGEHSHKAV-LARSKLES--LCRELQHRNLSKEE----- 242

Db 556 KYNQLQRLQDETALLRTESDTAARLKRTQABSSKQIQLESNNRDLQDNCLLETAKLK 615

Qy 243 -----GVQRAEEREEKKEVTSHFQVTLNDIQ-----LQMEQH 275

Db 616 LRKEPTNLSQSALESERRDR---THGSEINDIQGRICGLEEDLXNGKILLAKVELEKROL 672

Qy 276 NERNSEKLRQE--NMELAEARKLIEQVELREHIDKVPKHDLOQOLVD-----A 323

Db 673 QSRFTLEKSEKSNMEIDMTYQLKVTIQOSLEQBEAE---HKATKARLADKNKIYESIEEA 728

Qy 324 KLOOQOEMLEK-ABERHQREK-DFLLKEAVE-----SQPMCELMKQOETHLKQ 369

Db 729 KSEAMKMEKLEKLLBERTLKQKVENLLLEAKGCSLLDDCLKQSQKINELLKOKOV-LNE 787

Qy 370 QIALVTEKEEE-----PONTLSKSSSEVFTT-----FKQEMEKMTKKIKLEKETTM 415

Db 788 DVRNLTLEKTEQSTQKCLTQNDLKMTQVQNTLKMSEKQKQENNHLEMEKXNLEKQNAE 847

Qy 416 YSRVSESSNKALLEMAE-----KTVPRDKELEGLOVKIQRLKLCRALQTERNDLNKR 468

Db 848 LRKERQDADGQKQELQDQLEAEQYFTSTLYKTQVRLEKEECBEKTKLGLKQLOKQBLQDE 907

Qy 469 VQDLSAGGGSGLTDSGPER 487

Db 908 RDSLAQLEITLTKADSEQ 926

RESULT 47

US-08-685-576-4

; Sequence 4, Application US/08685576

; Patent No. 5906819

; GENERAL INFORMATION:

; APPLICANT: Kaibuchi, Kozo

; APPLICANT: Iwamatsu, Akihiro

; APPLICANT: Nakano, Takeshi

; APPLICANT: Ito, Masaaki

; APPLICANT: Takahashi, No. 5906819uaki

; TITLE OF INVENTION: RHO TARGET PROTEIN RHO-KINASE

; NUMBER OF SEQUENCES: 16

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Foley & Lardner

; STREET: 3000 K Street, N.W., Suite 500

; CITY: Washington

; STATE: D.C.

; COUNTRY: USA

; ZIP: 20007-5109

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; COMPUTER: IBM PC compatible

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: Patent In Release #1.0, Version #1.30

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/08/685,576

; FILING DATE: 24-JUL-1996

; CLASSIFICATION: 435

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: JP 7-325129

; FILING DATE: 20-NOV-1995

; APPLICATION DATA:

; APPLICATION NUMBER: JP 8-17150

; FILING DATE: 05-JAN-1996

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: JP 8-131206

; FILING DATE: 26-APR-1996

; ATTORNEY/AGENT INFORMATION:

; NAME: Best, Stephen A.

; REGISTRATION NUMBER: 29,768

; REFERENCE/DOCKET NUMBER: 16887/843

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: (202)672-5300

; TELEFAX: (202)672-5399

; TELEX: 904136

; INFORMATION FOR SEQ ID NO: 4:

```

; SEQUENCE CHARACTERISTICS:
; LENGTH: 198 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-685-576-4

Query Match          7.9%; Score 214; DB 2; Length 1388;
Best Local Similarity 20.5%; Pred. No. 2.8e-06;
Matches 127; Conservative 87; Mismatches 175; Indels 230; Gaps 21;

Qy   91 PEDAEKR-----TVVANGSPE-----PPVV---113
      |||||:::
Db   316 PDAEISKHAKNLCIFLTDREVGLGRNGVEIRQHFFFKNDQWHDNRITETAAFPVPVL 375
      :|||:||||
Qy   114 -YGEKPSKGDPTSEI--RQSDVG 136
      ::|||:||||
Db   376 SSDIDSNFDDIEDKDGVTFPFKPAPGVLPIGFTTYREMLLLSDSPSCHEDSIQ 435
      :|||:||||
Qy   137 DRHRRQE-----KKAKGLGKBKITLLMOTL 163
      :|||:||||
Db   436 SRNEESQEIQKKLYTLEHLNEMQAEELEQCKSVNTRLEXTAKELBEEITLRKSVE 495
      :|||:||||
Qy   164 NTLSTPEKLAALCKVAE-----LLEHRNSQKMK 195
      :|||:||||
Db   496 SALRQLEREXALLQHKAAYQRKADHEADKENLDVNSLKQLEDLKKGNQNSQISTE 555
      :|||:||||
Qy   196 LQKKQSQLVQEKDHLRGESHKAV-LARSKLES--LCRELQRHNRSLKEE-----242
      :|||:||||
Db   556 KVNQLRQLDSTNALRLTESDTAARLRKTAQAESKQIQOLESNNRDLDKNCILLETAKLK 615
      :|||:||||
Qy   243 -----GVQPARDEEEKRKVTSHFQVTLNDIQ-----LQMEQH 275
      :|||:||||
Db   616 LEKEFINLQSALESERDR---THGSEIINDLQRIGLSEDLGKGKILLAKVELEKQL 672
      :|||:||||
Qy   276 NERNSKURQE--NWELAEURLKLHQVELAREHIDKVFKHKDLQQQLVD-----A 323
      :|||:||||
Db   673 QERFTDLEKSKSNBIIDMTYQLKVIQOSLEGESAE--HKATKARLADKNNKIYESIEEA 728
      :|||:||||
Qy   324 KLQQAQBMIXE-AEERHOREK-DFLLEAVE-----SQRMCELMKQOEITLKO 369
      :|||:||||
Db   729 KSEAMKEMKXLLJEERTLKQVENLLLEAKRCSLDCDLKQSOOKINELLKQDV-LNE 787
      :|||:||||
Qy   370 QLALYTEKPE-----PONTLSKSEVFTT-----PKQENRYTKIKLKEKETTMM 415
      :|||:||||
Db   788 DVRLNLTKIPQTQKRCLTQNDLKMQTQOVNTLKMSKQLKQENNHLMKMNMLEKONAE 847
      :|||:||||
Qy   416 YRSRWSSNKALLEMASE-----KTVRKLEGLQVKIORLEKCRALQTERNDLNKR 468
      :|||:||||
Db   848 LRKERQADQGKMLQOLEAQVFSLYKTQVRELKEEBEKTCLGKELQKKQELQDE 907
      :|||:||||
Qy   469 VQDSLACGGSLTDSGPFR 487
      :|||:||||
Db   908 RDSLAAQLEILTAKADSEQ 926
      :|||:||||

RESULT 48
US-08-714-741-40
; Sequence 40, Application US/08714741
; Patent No. 650613
; GENERAL INFORMATION:
; APPLICANT: Briles, David E.
; APPLICANT: McDaniel, Larry S.
; APPLICANT: Swiatlo, Edwin
; APPLICANT: Yoher, Janet
; APPLICANT: Crain, Marilyn J.
; APPLICANT: Hollingshead, Susan
; APPLICANT: Tart, Rebecca
; APPLICANT: Brooks-Walter, Alexis
; TITLE OF INVENTION: ENEMGOCCAL GENES, PORTIONS THEREOF,
; EXPRESSION PRODUCTS THEREFROM, AND USES OF SUCH GENES
; TITLE OF INVENTION: EXPRESSION PRODUCTS THEREFROM, AND USES OF SUCH GENES
; NUMBER OF SEQUENCES: 47
```

```

RESULT 49
US-09-150-867-1
; Sequence 1, Application US/09150867
; Patent No. 6645748
; GENERAL INFORMATION:
; APPLICANT: Wood, Kenneth W.
; APPLICANT: Sakowicz, Roman
; APPLICANT: Goldstein, Lawrence S.B.
; APPLICANT: Cleveland, Don W.
; APPLICANT: The Regents of the University of California
; TITLE OF INVENTION: Plus End-Directed Microtubule Motor Required for
; TITLE OF INVENTION: Chromosome Congression
; FILE REFERENCE: 18557C-000110US
; CURRENT APPLICATION NUMBER: US/09/150,867
; EARLIER FILING DATE: 1998-09-10
; EARLIER APPLICATION NUMBER: US 60/059,645
; EARLIER FILING DATE: 1997-09-11
; NUMBER OF SEQ ID NOS: 11
; SOFTWARE: Patentin Ver. 2.0
; SEQ ID NO 1
; LENGTH: 2954
; TYPE: PRT
; ORGANISM: Xenopus sp.
; NAME/KEY: DOMAIN
; LOCATION: (1)..(472)
; OTHER INFORMATION: kinesin like motor domain
; FEATURE:
; NAME/KEY: DOMAIN
; LOCATION: (473)..(2752)
; OTHER INFORMATION: rod domain
; FEATURE:
; NAME/KEY: DOMAIN
; LOCATION: (2753)..(2954)
; OTHER INFORMATION: tail domain
; US-09-150-867-1

Query Match 7.8%; Score 212; DB 4; Length 2954;
Best Local Similarity 22.2%; Pred. No. 9.9e-06;
Matches 127; Conservative 83; Mismatches 169; Indels 192; Gaps 25;

QY 52 LRDVSEL--SRQLEILSTYCVANNQ--GPGDGAQGEPAPEPAEKSRITYVARNGEP 107
DB 1284 LRAAQBELREQQLVDSFRQQLDCSVGISPNHD----- 1318

QY 108 EPTPVYVGEKPSKGPDPTEBIRQSEV--GDRDHRPQEK----- 146
DB 1319 ----AVANQKYSVSGVNS--LQSEMLRGERDELQTSCKALVSELELLRAHVKSVEGEN 1371

QY 147 ----KAKGIGKEITLLMOTLNTLSTPBKLAALCKYABILLSEHRNSQKMKLIQ--KK 200
DB 1372 LEITKGLNGLEKEILGKSESEVYLSKMLNLEDNNKLEQAEYSSKENQPSLEEVFSG 1431

QY 201 QSOLOVEKDLRGEHSKAVIARSLKESLRE--LQRENRSL-----KEEG 243
DB 1432 SQKLVDEIEVLKAO--LKA--AEERLEIKDRDFELVOTANTVWEGKLETPLOADHEEDS 1488

QY 244 VQAREBEK--KEYTSHFQVTLNDIQLOMQHNRNSKLIR--QENME-----LAB 291
DB 1489 IDRRSEMEIKVLGKELRNQVLLERLQ--BEKLELSNKLRLQKEMETSVLLKDDLQ 1545

QY 292 RLKLLIQEYELREHDKVFK--HKDLQQQLVDKALQQAQEMLKE----- 334
DB 1546 KLESLSSENIILKENITDKHSOTQAQL--QKTOQELQLANLAIASDNCNPTQKE 1603

QY 335 -----AEERHQ-------EKDPLKAEVRSOR 355
DB 1604 TSADCVHPELEKILLUTBELHOKTNEQKLLHEKNELEQAQVELKCEVHELMKSMIESKS 1663

356 MCELMKQOETH-----LQOALALYTEKEPEEFONT-----LSKSEVFTTEKQ 397
DB 1664 SLESLL-QHEKHDTQOALLAQOQVQVTOBKKELOQTHEHLTAQVVDHLKENIELGLNFKN 1722

QY 398 EMERKTK-----IKKLEKETTMYRSRWSSKAL-----LEMAEKTVR-DKE 440
DB 1723 EAQKTTKBOCLLNENKELQSOHRLQCEIBELMKSLADKESALETLKESBQKVINLQOE 1782

QY 441 LEGQVKIORLEKLCRALQOTERNDLNKRQVD 471
DB 1783 MEMVMEBELKNSORTVIAERDQLODLRE 1813

RESULT 50
US-09-914-259-27
; Sequence 27, Application US/09914259
; Patent No. 6495336
; GENERAL INFORMATION:
; APPLICANT: Makowski, Lee
; APPLICANT: Hyman, Paul
; APPLICANT: Williams, Mark
; TITLE OF INVENTION: STAGED ASSEMBLY OF NANOSTRUCTURES
; FILE REFERENCE: 8471-010-999
; CURRENT APPLICATION NUMBER: US/09/914,259
; CURRENT FILING DATE: 2000-11-21
; NUMBER OF SEQ ID NOS: 180
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 27
; LENGTH: 1027
; TYPE: PRT
; ORGANISM: Mus musculus
; US-09-914-259-27

Query Match 7.8%; Score 211.5; DB 4; Length 1027;
Best Local Similarity 23.4%; Pred. No. 2.9e-06;
Matches 117; Conservative 87; Mismatches 223; Indels 73; Gaps 17;

QY 47 AOSGALRDVSELRLQLEEDILSTYCVNNQGGPGDGAQGEPAPEDAEKSRITYVARNGE 106
DB 475 SENDAAKDEVKEVLQALEELAVNY--DQKSEVTEKSOQQLLVDELQKVTATMLSELE 532

QY 107 PEPTPVYVGEKPSKGPDPTEBIRQSDS----VGDRDHRPQEKKAKGLKGIITLMTQ 162
DB 533 LQRLQVSGHQKRIAEVLNGLMRDLSEFVIVGNGBIKLPVEISGA--IBEEFTVARLY 590

QY 163 LNTLSTPEKLAALKKYAEIL--LEEHRNSQKMKLIQKQSQLYQCKDHLRG--BHSKAV 219
DB 591 ISKIKSEVKSVMKRCQLENLQVECHRRQKQVETGRELSSCQLLISQHEAKISLHEIMQTV 650

QY 220 -LARSKE-----SLCRSELQ-----HNRSLKEEGVQARAEERKEKKEVTSHFQVTLND 267
DB 651 ELKKRHLEESYDLSDELARLQAHETVHEVALKDKEPDTQDAEEVKA----- 698

QY 268 IQLOMQHNRNSKLIRQENMELARLKLIEQYELREHIDKVPFKDLOQ--QLVDAKL 325
DB 699 LELQENHREAHHR-----QLARLDRINEKQKTDIDEL---KDLNOKLQLELEKL 745

QY 326 QOAEMLKEAE-----ERFOREKDFL--LKEAV--ESORMCBLMKQOETHL 367
DB 746 QADYERLKNENEKSAKQELTFLYERHQSDQLKGLBETVARELQTLHLNRLKLFVQDV 805

QY 368 KOQALYTEKEPEEFONTLSKSESVFTTFKQEMERKTKIKKLEKETMYRSRWSSKAL 427
DB 806 TTRVKSAAEMEPEDSGGTHSQKISFLENNLEQLTKVHKQLVRDNLRCLEPKLEKL 865

QY 428 LEMAEKTVRKDEGL--LOVKIORLEKLCRALQOTERNDLNKRVDLSAGGOSLTDSCPE 486
DB 866 RATAE-----RVKALEGALKKEAKGAKDKRKYQOEVDRIKAVRYKSSGKSGHSAQIAKP 921

QY 487 RRPQGAQAPSSPRVTEAP 506
DB 922 VRPGHYPASSPTNPYGRSP 941

```

Search completed: June 7, 2004, 14:42:05
Job time : 40 secs

107	7	1.3	1064	1	CARB_BACST	OS0302 bacillus st	180	1.1	123	1	PFDE_AERPE	O9yc11 aeropyrum p
108	7	1.3	1065	1	CARB_BACCL	P46537 bacillus ca	181	1.1	126	1	RS11_METAC	O8trr0 methanosarc
109	7	1.3	1071	1	DAMI_MOUSE	Q8bpm0 mus musculu	182	1.1	126	1	RS11_METMA	O8pv17 methanosarc
110	7	1.3	1078	1	DAMI_HUMAN	O9v4d1 homo sapien	183	1.1	126	1	SYN2_MOUSE	O64332 mus musculu
111	7	1.3	1079	1	IR2P_SCHPO	Q10251 schizosacch	184	1.1	127	1	MINE_CHLVU	P56350 chlorella v
112	7	1.3	1085	1	YAPA_SCHPO	Q09863 schizosacch	185	1.1	127	1	NB7M_BOVIN	O26387 bos taurus
113	7	1.3	1139	1	SRE2_CRIGR	Q60429 cricetulus	186	1.1	127	1	RS11_HALN1	Q9hgj5 halobacteri
114	7	1.3	1141	1	SRE2_HUMAN	Q12772 homo sapien	187	1.1	128	1	RR9_ASTIO	P58135 astasia lon
115	7	1.3	1159	1	SOR2_HUMAN	Q96pqc homo sapien	188	1.1	128	1	RS11_HALMA	P10788 haloarcula
116	7	1.3	1191	1	TOP2_ASFZ2	P34203 african swi	189	1.1	129	1	RS11_METJA	P54021 methanococc
117	7	1.3	1192	1	TOP2_ASFZ7	Q00942 african swi	190	1.1	130	1	RS11_METTH	O26143 methanobact
118	7	1.3	1372	1	RPOC_RICCN	O9th40 rickettsia	191	1.1	130	1	RS11_THEAC	O9hjd8 thermoplasm
119	7	1.3	1372	1	RPOC_RICCP	O9ze20 rickettsia	192	1.1	130	1	VG67_BPH2	P16517 bacterioph
120	7	1.3	1453	1	Y373_BOVIN	Q9cu23 bos taurus	193	1.1	131	1	CAIF_ECOLI	O47081 escherichia
121	7	1.3	1498	1	G0A3_HUMAN	Q08378 homo sapien	194	1.1	131	1	GC5H_XANCP	Q8p138 xanthomonas
122	7	1.3	1657	1	G0A3_HUMAN	P46940 homo sapien	195	1.1	131	1	GC5H_XANCP	O8p6c9 xanthomonas
123	7	1.3	1657	1	IQG1_MOUSE	Q9jklf1 mus musculu	196	1.1	131	1	RS11_AERPE	O9yb55 aeropyrum p
124	7	1.3	1940	1	IVH3_HUMAN	P11055 homo sapien	197	1.1	132	1	CYCP_PARSF	F00143 paracoccus
125	7	1.3	1940	1	IVH3_RAT	P12847 rattus norv	198	1.1	132	1	PPDA_PYRAE	O8tct9 pyrobaculum
126	7	1.3	2144	1	BP28_HUMAN	O9b583 homo sapien	199	1.1	132	1	RS11_THEVO	Q97b94 thermoplasm
127	7	1.3	3034	1	CLR1_MOUSE	O35161 mus musculu	200	1.1	133	1	NL21_PARJU	P55958 parietaria
128	7	1.3	4568	1	DH8B_CHLRE	Q39565 chlamydomon	201	1.1	133	1	NL21_PARJU	O04403 parietaria
129	7	1.3	5171	1	BPEA_HUMAN	O94833 homo sapien	202	1.1	133	1	RS11_ARCFU	O28001 archaeoglob
130	7	1.3	7389	1	BPAL_MOUSE	Q91zu6 mus musculu	203	1.1	134	1	ACP1_BRANA	P10352 brassica na
131	7	1.3	8797	1	SNEL_HUMAN	Q8nf91 homo sapien	204	1.1	134	1	ACP2_BRANA	P17650 brassica na
132	6	1.1	49	1	LHB2_RHOCA	P07368 rhodobacter	205	1.1	134	1	ACP3_BRANA	P32897 brassica na
133	6	1.1	56	1	YDAE_ECOLI	P38394 escherichia	206	1.1	134	1	ACP_BACRM	P07088 brassica ca
134	6	1.1	66	1	SCXL_ANDAU	P80950 androctonus	207	1.1	134	1	ATPE_FUSNN	O8rga3 fusobacteri
135	6	1.1	71	1	Y0TK_FMV	P20954 papaya mosa	208	1.1	135	1	YI04_MRIJA	Q8303 methanococc
136	6	1.1	71	1	EX7S_STRAS	Q866m0 streptococc	209	1.1	135	1	ACP2_ARATH	P25701 arabidopsis
137	6	1.1	75	1	R17E_THEVO	Q97b13 thermoplasm	210	1.1	136	1	ACP3_ARATH	P25702 arabidopsis
138	6	1.1	76	1	RPOZ_GUITH	O78442 guillardia	211	1.1	136	1	Y914_METJA	O58324 methanococc
139	6	1.1	77	1	RL29_MYCBO	O60650 mycobacteri	212	1.1	137	1	RS11_METKA	O8tvt9 methanopyru
140	6	1.1	77	1	RL29_MYCCTO	P95057 mycobacteri	213	1.1	137	1	RS11_PYRFO	O8u0a3 pyrococcus
141	6	1.1	79	1	EX7S_ECOLI	P22938 escherichia	214	1.1	137	1	RS11_PYRHO	O59304 pyrococcus
142	6	1.1	79	1	EX7S_SALTY	Q8xez9 salmonella	215	1.1	138	1	NUSB_PHOLL	O7n0j0 photorhabdu
143	6	1.1	79	1	TATA_CAMJE	Q9pnb9 campylobact	216	1.1	138	1	NUSB_YERPE	Q6zc42 yersinia pe
144	6	1.1	80	1	AKA7_MOUSE	O55074 mus musculu	217	1.1	139	1	RS14_SCHPO	O14150 schizosacch
145	6	1.1	83	1	RS27_HUMAN	P42677 homo sapien	218	1.1	141	1	HEA2_TACAC	F01978 tachygllossu
146	6	1.1	83	1	RS27_XENLA	P47904 xenopus lae	219	1.1	141	1	RS12_PLAFA	O97249 plasmodium
147	6	1.1	83	1	VEB3_AGR75	P08061 agrobacteri	220	1.1	141	1	TCL6_HUMAN	P56846 homo sapien
148	6	1.1	83	1	YN78_RHIME	Q9x7l3 rhizobium m	221	1.1	141	1	XKDS_BACSU	P54338 bacillus su
149	6	1.1	89	1	CH10_PORGI	P42376 porphyromon	222	1.1	141	1	ZRAP_ECOLI	P32682 escherichia
150	6	1.1	91	1	RR19_PEA	P31162 pisum sativ	223	1.1	144	1	RS14_TRYBB	P19800 trypanosoma
151	6	1.1	91	1	YAIN_ECOLI	P57556 escherichia	224	1.1	144	1	SODM_ARCFU	P28766 parastichop
152	6	1.1	92	1	Y247_CHLCV	Q824a6 chlamydophi	225	1.1	144	1	Y047_ARCFU	O30189 archaeoglob
153	6	1.1	93	1	GATC_HELPY	Q92km6 helicobacte	226	1.1	145	1	THI3_CORNE	P52228 corynebacte
154	6	1.1	93	1	GATC_HELPY	O25626 helicobacte	227	1.1	145	1	YIAA_ECOLI	P11297 escherichia
155	6	1.1	95	1	EMGY_MOUSE	P17095 mus musculu	228	1.1	146	1	CDD_MOUSE	P56389 mus musculu
156	6	1.1	96	1	Y143_BORBU	O51168 borrelia bu	229	1.1	147	1	YRAK_BACSU	P50862 bacillus su
157	6	1.1	96	1	Y157_METUA	Q57621 methanococc	230	1.1	148	1	DUT_RICCN	Q92174 rickettsia
158	6	1.1	99	1	GATC_FALSO	Q8y3c2 raietonia s	231	1.1	148	1	DUT_RICPR	P27151 rickettsia
159	6	1.1	99	1	Y088_UREPA	Q9pr57 ureaplasma	232	1.1	148	1	RL9_THETH	P17837 pseudomonas
160	6	1.1	100	1	Y667_CHLMU	Q08648 homo sapien	233	1.1	150	1	PMCD_PSEAE	P17837 pseudomonas
161	6	1.1	103	1	SPGB_HUMAN	Q8648 homo sapien	234	1.1	150	1	PDAD_METKA	O8txd4 methanopyru
162	6	1.1	103	1	SPGB_PANTR	Q8mz28 pan troglod	235	1.1	150	1	YVBA_BACSU	P37503 bacillus su
163	6	1.1	105	1	TYM4_VACGS	P33560 macrothete	236	1.1	152	1	ADOM_BPT3	P07693 bacterioph
164	6	1.1	105	1	Y0S2_YEAST	P46986 saccharomyc	237	1.1	152	1	MGSA_SALTY	Q8xer5 salmonella
165	6	1.1	106	1	HNGI_CRIGR	Q9gxp3 cricetulus	238	1.1	153	1	HOXT_AZOVI	P30781 azotobacter
166	6	1.1	106	1	HNGI_HUMAN	P17096 homo sapien	239	1.1	153	1	YE34_COXBU	Q83br9 coxiella bu
167	6	1.1	106	1	YTH5_RHOER	P43495 rhodococcus	240	1.1	154	1	Y712_RICPR	O9zcl5 rickettsia
168	6	1.1	109	1	CDK1_MOUSE	O35207 mus musculu	241	1.1	155	1	RR7_GUSEU	P46292 cuscutea eur
169	6	1.1	109	1	RLA1_MAI2E	P52855 zea mays (m	242	1.1	155	1	YHBI_HCMVA	P09702 human cytom
170	6	1.1	109	1	VHEM_PVMR	P17527 potato viru	243	1.1	157	1	Y53C_AGR15	O8uhx2 agrobacteri
171	6	1.1	110	1	YHBJ_ACTAC	P96769 actinobacil	244	1.1	157	1	Y93S_AQAAE	Q67076 aquifex ae
172	6	1.1	112	1	Y0DB_BACSU	O34844 bacillus su	245	1.1	158	1	Y612_RHIME	Q92s95 rhizobium m
173	6	1.1	113	1	PTWX_ECOLI	P32676 escherichia	246	1.1	159	1	FMF5_ECOLI	P13719 escherichia
174	6	1.1	114	1	CDK1_MESAU	P49119 mesocricetu	247	1.1	163	1	DTD_NEIMB	O9k143 neisseria m
175	6	1.1	115	1	CALR_PHYBI	P81564 phyllomedus	248	1.1	165	1	MSR2_UREPA	O9pqk2 ureaplasma
176	6	1.1	115	1	CDK1_HUMAN	O14519 homo sapien	249	1.1	165	1	RS16_CAUCR	P58122 caulobacter
177	6	1.1	117	1	GNPK_HALME	Q02236 halobacteri	250	1.1	165	1	RUVX_RHIME	Q92ql0 rhizobium m
178	6	1.1	118	1	YNIF_AZOER	P35316 azospirillu	251	1.1	166	1	NHAX_BACSU	O07552 bacillus su
179	6	1.1	121	1	Y879_HALN1	Q9hr37 halobacteri	252	1.1	166	1	VDEL_BPP4	P12551 bacterioph

253	6	1.1	167	1	FIXE RHIME	P18397 rhizobium m	RT
254	6	1.1	167	1	Y491 METJA	Q57914 methanococ	RT
255	6	1.1	168	1	Y224 HABIN	P44580 haemophilus	RL
256	6	1.1	168	1	YD02 METJA	Q58698 methanococ	RN
257	6	1.1	169	1	DSB2_PSEAE	P57701 pseudomonas	RP
258	6	1.1	170	1	RIMM_XANCP	Q8pbc2 xanthomonas	RX
259	6	1.1	170	1	RIMM_XYLFA	Q9ph38 xylella fas	RA
260	6	1.1	170	1	RIMM_XYLFA	Q8f155 xylella fas	RA
261	6	1.1	171	1	YF87 METJA	Q58982 methanococ	RL
262	6	1.1	171	1	ADXH DROME	P37193 drosophila	RN
263	6	1.1	172	1	IF3 CAUCR	Q9a9d9 caulobacter	RP
264	6	1.1	173	1	DEF RHIME	Q92sh6 rhizobium m	RC
265	6	1.1	174	1	RPCI_BPMU	P06019 bacterioph	RX
266	6	1.1	175	1	AXIK_ARATH	C24410 arabidopsis	RA
267	6	1.1	175	1	BADR_RHOPA	O07458 rhodospheu	RA
268	6	1.1	175	1	SSB SERMA	P25762 serratia ma	RA
269	6	1.1	177	1	NUOE RICPR	Q9zdh5 rickettsia	RA
270	6	1.1	178	1	NI2M HUMAN	Q9y6m9 homo sapien	RA
271	6	1.1	178	1	PYRR_ENTFA	O52707 enterococu	RA
272	6	1.1	179	1	GRPE_LACIG	Q9cgv9 lactococcus	RA
273	6	1.1	179	1	GRPE_LACIG	P42369 lactococcus	RA
274	6	1.1	179	1	GRPE_STRMU	O06941 streptococ	RA
275	6	1.1	179	1	RIMM_MYCLE	O33016 mycobacteri	RA
276	6	1.1	179	1	YMDB SALTY	Q9zq29 salmonella	RA
277	6	1.1	180	1	KDOP_HABIN	P45314 haemophilus	RA
278	6	1.1	180	1	YS76_AQUAE	O67453 aquifex aeo	RA
279	6	1.1	181	1	HE28_HUMAN	Q13442 homo sapien	RA
280	6	1.1	182	1	ARL3_HUMAN	P36405 homo sapien	RA
281	6	1.1	183	1	ASPH_AERHY	P31735 aeromonas h	RA
282	6	1.1	184	1	AROK_CHLMU	Q9pk27 chlamydia n	RA
283	6	1.1	184	1	IPYR_MYCFN	P47593 mycoplasma	RT
284	6	1.1	184	1	IPYR_MYCFN	P75250 mycoplasma	RT
285	6	1.1	185	1	RFP_THETN	O8rae2 thermocoe	RA
286	6	1.1	185	1	RFP_HELHP	Q7v1i0 helicobacte	RA
287	6	1.1	185	1	YHQO_YEAST	P38812 saccharomyc	CC
288	6	1.1	186	1	ATPD_CYAPA	P48082 cyanophora	CC
289	6	1.1	186	1	ATPF_MESVI	Q9mut1 mesostigma	CC
290	6	1.1	186	1	COAE_TETH	O56416 thermus the	CC
291	6	1.1	186	1	RRF_BABA	O8rt64 bartonella	CC
292	6	1.1	186	1	YRKQ_BACSU	P54430 bacillus su	CC
293	6	1.1	187	1	Y4GA_RHLSN	P55457 rhizobium s	CC
294	6	1.1	189	1	R13A_SALTR	Q91487 salmo trutt	CC
295	6	1.1	189	1	RK12_SPIOL	P02398 spinacia ol	CC
296	6	1.1	190	1	GRPE_STRPY	Q99yc8 streptococ	CC
297	6	1.1	192	1	CDRC_CLOAB	P45812 clatridium	CC
298	6	1.1	192	1	INAI_RAT	P05011 rattus norv	CC
299	6	1.1	194	1	DMOI_DESMO	P21505 desulfuroco	CC
300	6	1.1	195	1	PYRE_SULSO	Q9ux09 sulfolobus	CC

ALIGNMENTS

RESULT 1	YL14_HUMAN	STANDARD;	PRT;	259 AA.
YL14_HUMAN	AC	P40222; Q8NRY3;		
AC	DT	01-FEB-1995 (Rel. 31, Created)		
DT	DT	28-FEB-2003 (Rel. 41, Last sequence update)		
DT	DT	10-OCT-2003 (Rel. 42, Last annotation update)		
DE	DE	Hypothetical protein initially thought to be identical with		
DE	DE	interleukin-14 (IL-14) (High molecular weight B-cell growth factor)		
DE	DE	(HMM-BCGF).		
GN	GN	IL14.		
GN	GN	Homo sapiens (Human).		
OS	OS	Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;		
OC	OC	Mammalia; Euthera; Primates; Catarrhini; Hominidae; Homo.		
OX	OX	NCBI_TaxID=9606;		
RN	RN	[1]		
RP	RP	SEQUENCE FROM N.A.		
RX	RX	MEDLINE=9337675; PubMed=8327514;		
RA	RA	Ambrus J.L. Jr., Pippin J., Joseph A., Xu C., Blumenthal D.,		
RA	RA	Tamayo A., Claypool K., McCourt D., Srikiatchachorn A., Ford R.J.;		

RT	*Identification of a cDNA for a human high-molecular-weight B-cell growth factor.*
RL	Proc. Natl. Acad. Sci. U.S.A. 90:6330-6334(1993).
RN	[2]
RP	ERRATUM
RX	MEDLINE=96353961; PubMed=8755619;
RA	Ambrus J.L. Jr., Pippin J., Joseph A., Xu C., Blumenthal D.,
RA	Tamayo A., Claypool K., McCourt D., Srikiatchachorn A., Ford R.J.;
RA	Proc. Natl. Acad. Sci. U.S.A. 93:8154-8154(1996).
RN	[3]
RP	SEQUENCE OF 130-259 FROM N.A.
RC	TISSUE=Placenta;
RX	MEDLINE=22388257; PubMed=12477932;
RA	Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA	Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
RA	Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA	Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
RA	Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA	Stapleton M., Soares M.B., Bonaldo M.P., Casavant T.L., Scheetz T.E.,
RA	Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullahy S.J.,
RA	Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA	Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
RA	Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA	Fahey J., Helton E., Kettman M., Madan A., Rodrigues S., Sanchez A.,
RA	Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA	Blakesley R.W., Touchman J.W., Green B.D., Dickson M.C.,
RA	Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,
RA	Butterfield Y.S.N., Krzywinski M.I., Skalska U., Smallos D.E.,
RA	Schmerch A., Schein J.E., Jones S.J.M., Marra M.A.;
RT	"Generation and initial analysis of more than 15,000 full-length
RT	human and mouse cDNA sequences."
RL	Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
CC	!- CAUTION: The clone described in Ref.1 was initially thought to
CC	code for a high molecular weight interleukin (IL-14). The
CC	translation of this cDNA does not predict the open reading frame
CC	for the 60 kDa protein described in Ref.1. A reading frame on the
CC	plus strand predicts a 7.7 kDa protein. The longest open reading
CC	frame (shown here) is on the opposite strand and predicts a 36.4
CC	kDa protein. The relationship of this sequence to IL-14, if any,
CC	is uncertain.
CC	-----
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration
CC	between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC	European Bioinformatics Institute. There are no restrictions on its
CC	use by non-profit institutions as long as its content is in no way
CC	modified and this statement is not removed. Usage by and for commercial
CC	entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC	or send an email to license@isb-sib.ch).
CC	-----
DR	EMBL; LI5344; -; NOT ANNOTATED CDS.
DR	EMBL; BC029686; AAH29686.1; -;
DR	PIR; A48203; A48203.
DR	Genew; HGNC:5976; IL14.
DR	MIM; 147684; -;
DR	GO; GO:0005576; C:extracellular; NAS.
DR	GO; GO:0003072; F:high molecular weight B-cell growth factor . . .; NAS.
DR	GO; GO:0008283; P:cell proliferation; NAS.
DR	Hypothetical protein.
SW	SEQUENCE 259 AA; 30107 MW; 8AB8D1AA6891357C CRC64;
SQ	
	Query Match 48.9%; Score 259; DB 1; Length 259;
	Best Local Similarity 100.0%; Pred. No. 3.6e-250; Indels 0; Gaps 0;
	Matches 259; Conservative 0; Mismatches 0;
QY	272 MEQHNRNSKLRQENNELAERLKKLIEQYELREEHIDKVFKEHDLQQLVDAKLAQAQRM 331
DB	1 MEQHNRNSKLRQENNELAERLKKLIEQYELREEHIDKVFKEHDLQQLVDAKLAQAQRM 60
QY	332 LKEAERHQREKDFLLKEAVSSQRMCELMKQOETHLKKOALYTKFEPFQNTLSKSSV 391
DB	61 LKEAERHQREKDFLLKEAVSSQRMCELMKQOETHLKKOALYTKFEPFQNTLSKSSV 120


```
QY 392 FTTFQEMKMTKKIKLEKFTMYRSRWSSNKALLEMAEKTVRDKLEGLQVKIQL 451
DB 121 FTTFQEMKMTKKIKLEKFTMYRSRWSSNKALLEMAEKTVRDKLEGLQVKIQL 180
QY 452 EKLCEALQTERNDLNKRVDLSAGCGSITDSGPERDEGCAOPSPRVTEAPCPYGA 511
DB 181 EKLCEALQTERNDLNKRVDLSAGCGSITDSGPERDEGCAOPSPRVTEAPCPYGA 240
QY 512 PSTEASGOTGPQEPPTSARA 530
DB 241 PSTEASGOTGPQEPPTSARA 259

RESULT 2
ID FLO8_YEAST STANDARD; PRT; 799 AA.
AC P40067; Q05751;
DT 01-FEB-1995 (Rel. 31, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Transcriptional activator FLO8 (PH5 protein).
GN FLO8 OR PH5 OR YER105C/YER108C.
OS Saccharomyces cerevisiae (Baker's yeast).
OC Sukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
OX NCBI_TaxID=4932;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Sigma 12789;
RA Liu H., Styles C.A., Fink G.;
RL Submitted (APR-1996) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE OF 71-799 FROM N.A.
RC STRAIN=ATCC 60715;
RX MEDLINE=96335146; PubMed=8757402;
RA Kobayashi O., Suda H., Ohtani T., Sone H.;
RT "Molecular cloning and analysis of the dominant flocculation gene
  FLO8 from Saccharomyces cerevisiae.";
RL Mol. Gen. Genet. 251:707-715(1996).
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=S288C / AB972;
RX MEDLINE=97313264; PubMed=9169868;
RA Dietrich F.S., Mulligan J.T., Hennessy K.M., Yelton M.A., Allen E.,
RA Araujo R., Aviles E., Berno A., Brennan T., Carpenter J., Chen E.,
RA Cherry J.M., Chung E., Duncan M., Guzman E., Hartzell G.,
RA Hunnicke-Smith S., Hyman R.W., Kayser A., Komp C., Lashkari D., Lew H.,
RA Lin D., Mosedale D., Nakahara K., Namath A., Norgren R., Oefner P.,
RA Oh C., Petel P.X., Roberts D., Sehl P., Schramm S., Shogren T.,
RA Smith V., Taylor P., Wei Y., Botstein D., Davis R.W.;
RT "The nucleotide sequence of Saccharomyces cerevisiae chromosome V.";
RL Nature 387:78-81(1997).
CC -!- FUNCTION: Required for diploid filamentous growth, haploid
  invasive growth and flocculation. Putative transcriptional
  activator of FLO1.
CC -!- SUBCELLULAR LOCATION: Nuclear.
CC -!- SIMILARITY: Contains 1 Lish domain.
CC -!- CAUTION: Ref.3 sequence differs from that shown due to a
  frameshift.

This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (see http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).

EMBL; U51431; AAC49522.1; -
DB EMBL; D83713; BRA12076.1; -
DB EMBL; U18916; AAC03207.1; ALT_FRAME.
DB EMBL; U18916; AAC03206.1; ALT_FRAME.
DB GeneOnline; 139188; -

TRANSFAC; T03311; -.
SGD; S0000911; FLO8.
GO; GO:0005737; C:cytoplasm; IDA.
GO; GO:0005634; C:nucleus; IDA.
InterPro; IPR006594; Lish.
SMART; SM00667; Lish; 1.
PROSITE; PS00896; Lish; 1.
KW Transcription regulation; Activator; Nuclear protein.
FT DOMAIN 73 105 POLY-GLN.
FT DOMAIN 41 55
FT CONFLICT 112 112 V -> I (IN REF. 2).
FT CONFLICT 115 115 P -> S (IN REF. 2).
FT CONFLICT 383 383 C -> G (IN REF. 2).
FT CONFLICT 441 441 A -> T (IN REF. 2).
FT CONFLICT 447 447 A -> V (IN REF. 2).
FT CONFLICT 598 598 R -> P (IN REF. 2).
SQ SEQUENCE 799 AA; 86834 MW; 91AAE10D8E586DAF CRC64;

Query Match 1.7%; Score 9; DB 1; Length 799;
Best Local Similarity 100.0%; Pred. No. 1.8;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 164 NTLSTPEEK 172
DB 671 NTLSTPEEK 679

RESULT 3
CARB_METAC
ID - CARB_METAC STANDARD; PRT; 1070 AA.
AC Q8TNY4;
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Carbamoyl-phosphate synthase large chain (EC 6.3.5.5) (Carbamoyl-
  phosphate synthetase ammonia chain).
OS Carb OR MA2143.
OS Methanosarcina acetivorans.
OC Archaea; Euryarchaeota; Methanomicrobia; Methanosarcinales;
OC Methanosarcinaceae; Methanosarcina.
OX NCBI_TaxID=2214;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C2A / ATCC 35395 / DSM 2834;
RX MEDLINE=21929760; PubMed=11932238;
RA Galagan J.E., Nusbaum C., Roy A., Endrizzi M.G., MacDonald P.,
RA FitzHugh W.E., Calvo S., Engels R., Smirnov S., Atnoor D., Brown A.,
RA Allen N., Naylor J., Staiger-Thomann N., DeArelano K., Johnson R.,
RA Linton L., McEwan P., McKernan K., Talamas J., Tirrell A., Ye W.,
RA Zimmer A., Barber R.D., Cann I., Graham D.E., Grahame D.A., Guss A.M.,
RA Heddreich R., Ingram-Smith C., Kuettnner H.C., Krzycki J.A.,
RA Leigh J.A., Li W., Liu J., Mukhopadhyay B., Reeve J.N., Smith K.,
RA Springer T.A., Umayam L.A., White O., White R.H., de Macario E.C.,
RA Ferry J.G., Jarrell K.F., Jing H., Macario A.J.L., Paulsen I.,
RA Pritchett M., Sowers K.R., Swanson R.V., Zinder S.H., Lander E.,
RA Metcalf W.W., Birren B.;
RT "The genome of Methanosarcina acetivorans reveals extensive metabolic
  and physiological diversity.";
RL Genome Res. 12:532-542(2002).
CC -!- CATALYTIC ACTIVITY: 2 ATP + L-glutamine + CO(2) + H(2)O = 2 ADP +
  phosphate + L-glutamate + carbamoyl phosphate.
CC -!- COFACTOR: Binds 3 manganese ions per subunit (By similarity).
CC -!- PATHWAY: Arginine biosynthesis.
CC -!- PATHWAY: Pyrimidine biosynthesis; first step.
CC -!- SUBUNIT: Composed of two chains; the small (or glutamine) chain
  promotes the hydrolysis of glutamine to ammonia, which is used by
  the large (or ammonia) chain to synthesize carbamoyl phosphate (By
  similarity).
CC -!- SIMILARITY: Belongs to the carb family.

This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
```



```

OC Mammalia; Eutheria; Lagomorpha; Leporidae; Oryctolagus.
OX NCBI_TaxID=9986;
RN [1]
RP SEQUENCE FROM N.A., AND SEQUENCE OF 479-500.
RC STRAIN=New Zealand white; TISSUE=Brain;
RX MEDLINE=93374895; PubMed=8396129;
RA Lamphear B.J., Yan R., Yang P., Waters D., Liebig H.-D.,
RT Klump H., Kuechler E., Skern T., Rhoads R.E.;
RT "Mapping the cleavage site in protein synthesis initiation factor
RT eIF-4 gamma of the 2A proteases from human Coxsackievirus and
RT rhinovirus.";
RL J. Biol. Chem. 268:19200-19203(1993).
RN [2]
RP PARTIAL SEQUENCE.
RX MEDLINE=93054654; PubMed=1429670;
RA Yan R., Rychlik W., Etchison D., Rhoads R.E.;
RT "Amino acid sequence of the human protein synthesis initiation factor
RT eIF-4 gamma.";
RL J. Biol. Chem. 267:23226-23231(1992).
CC -!- FUNCTION: Component of the protein complex EIF4F, which is
CC involved in the recognition of the mRNA cap, ATP-dependent
CC unwinding of 5'-terminal secondary structure and recruitment of
CC mRNA to the ribosome.
CC -!- SUBUNIT: EIF4F is a multi-subunit complex, the composition of
CC which varies with external and internal environmental conditions.
CC It is composed of at least EIF4A, EIF4E and EIF4G. EIF4G interacts
CC with the serine/threonine kinases MNK1 and MNK2. Appears to act
CC as a scaffold protein, holding these enzymes in place to
CC phosphorylate EIF4E (By similarity).
CC -!- SIMILARITY: THE C-TERMINAL REGION IS SIMILAR TO THE N-TERMINAL
CC REGION OF WHEAT EURARYOTIC INITIATION FACTOR (ISO)4P SUBUNIT P82.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; L22090; AAA31242.1; -
DR PIR; I46707; I46707.
DR InterPro; IPR008938; ARM.
DR InterPro; IPR003307; eIF5C.
DR InterPro; IPR003890; IF_eIF4G.
DR InterPro; IPR003891; IF_eIF4G_MA3.
DR InterPro; IPR000504; RNA_rec_mot.
DR Pfam; PF02847; MA3; 1.
DR Pfam; PF02854; MIF4G; 1.
DR SMART; SM00515; eIF5C; 1.
DR SMART; SM00544; MA3; 1.
DR SMART; SM00543; MIF4G; 1.
KW Initiation factor; Protein biosynthesis; Translation regulation;
FT Phosphorylation; RNA-binding.
FT DOMAIN 188 192 POLY-PRO.
FT DOMAIN 262 275 POLY-GLU.
FT DOMAIN 1393 1398 POLY-GLU.
SQ SEQUENCE 1402 AA; 154050 MW; 7FD85D7E30519230 CRC64;
Query Match 1.7%; Score 9; DB 1; Length 1402;
Best Local Similarity 100.0%; Pred. No. 2.9;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 56 SEELSRQLE 64
DB 1243 SEELSRQLE 1251
RESULT 6
SYE DEIRA
ID_SYE DEIRA STANDARD; PRT; 483 AA.
AC Q9R330;

```

```

DT 10-OCT-2003 (Rel. 42, Created)
DT 10-OCT-2003 (Rel. 42, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Glutamyl-tRNA synthetase (EC 6.1.1.17) (Glutamate--tRNA ligase)
DE (Glurs).
DE GLTX OR DR0485.
OS Deinococcus radiodurans.
OC Bacteria; Deinococcus-Thermus; Deinococci; Deinococcales;
OC Deinococcaceae; Deinococcus.
OX NCBI_TaxID=1299;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN=R1 / ATCC 13939 / DSM 20539 / NCIB 9279;
RX MEDLINE=20036896; PubMed=10567266;
RA White O., Eisen J.A., Heidelberg J.F., Hickey E.K., Peterson J.D.,
RA Dodson R.J., Hatt D.H., Gwinn M.L., Nelson W.C., Richardson D.L.,
RA Moffat K.S., Qin H., Jiang L., Pamphile W., Crosby M., Shen M.,
RA Vamathevan J.J., Lam P., McDonald L., Utterback T., Zaleski C.,
RA Makarova K.S., Aravind L., Daly M.J., Minton K.W., Fleischmann R.D.,
RA Ketchum K.A., Nelson K.E., Salzberg S., Smith H.O., Venter J.C.,
RA Fraser C.M.;
RT "Genome sequence of the radioresistant bacterium Deinococcus
RT radiodurans R1.";
RL Science 286:1571-1577(1999).
CC -!- CATALYTIC ACTIVITY: ATP + L-glutamate + tRNA(Glu) = AMP +
CC diphosphate + L-glutamyl-tRNA(Glu).
CC -!- SUBUNIT: Monomer (By similarity).
CC -!- SUBCELLULAR LOCATION: Cytoplasmic.
CC -!- SIMILARITY: Belongs to class-I aminoacyl-tRNA synthetase family.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; A5001907; AAF10063.1; ALT_INIT.
DR PIR; H75514; H75514.
DR HSSP; P27000; 1GLN.
DR TIGR; DR0485; -.
DR HAMAP; MF_00022; -; 1.
DR InterPro; IPR004527; Glx_bact.
DR InterPro; IPR000924; Glu_tRNA-synt_1c.
DR InterPro; IPR008925; tRNA-synt_bind.
DR InterPro; IPR001412; tRNA-synt_1.
DR Pfam; PF00749; tRNA-synt_1c; 1.
DR PRINTS; PR00987; TRNASYNTHGLU.
DR TIGRPFAMS; TIGR00464; gltx_bact; 1.
DR PROSITE; PS00178; AA_TRNA_LIGASE_I; 1.
KW Aminoacyl-tRNA synthetase; Protein biosynthesis; Ligase; ATP-binding;
KW Complete proteome.
FT SITE 14 24 "HIGH" REGION.
FT SITE 253 257 "RMSKS" REGION.
FT BINDING 256 256 ATP (BY SIMILARITY).
SQ SEQUENCE 483 AA; 54689 MW; 7FG6F5C0BF3A4567 CRC64;
Query Match 1.5%; Score 8; DB 1; Length 483;
Best Local Similarity 100.0%; Pred. No. 11;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 496 AFSSPRVT 503
DB 3 AFSSPRVT 10
RESULT 7
MURD RICCN STANDARD; PRT; 500 AA.
ID MURD RICCN
AC Q92IE0;
DT 10-OCT-2003 (Rel. 42, Created)
DT 10-OCT-2003 (Rel. 42, Last sequence update)

```

```

DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE UDP-N-acetylmuramoyl-L-alanine--D-glutamate ligase (EC 6.3.2.9) (UDP-N-
DE acetylmuramoyl-L-alanyl-D-glutamate synthetase) (D-glutamic acid
DE adding enzyme).
GN MURD OR RC0560.
OS Rickettsia conorii.
OC Bacteria; Proteobacteria; Alphaproteobacteria; Rickettsiales;
OC Rickettsiaceae; Rickettsiae; Rickettsia.
OX NCBI_TaxID=781;
RN [1]
RN SEQUENCE FROM N.A.
RC STRAIN=Malish 7;
RX MEDLINE=21442074; PubMed=11557893;
RA Ogata H., Audic S., Renesto-Audiffren P., Fournier P.-E., Barbe V.,
RA Sanson D., Roux V., Cossart P., Weissbach J., Claverie J.-M.,
RA Raoult D.;
RT "Mechanisms of evolution in Rickettsia conorii and R. prowazekii.";
RL Science 293:2093-2098(2001).
CC -!- FUNCTION: Cell wall formation. Catalyzes the addition of glutamate
CC to the nucleotide precursor UDP-N-acetylmuramoyl-L-alanine (UMA)
CC (By similarity).
CC -!- CATALYTIC ACTIVITY: ATP + UDP-N-acetylmuramoyl-L-alanine +
CC Glutamate = ADP + phosphate + UDP-N-acetylmuramoyl-L-alanyl-D-
CC Glutamate.
CC -!- PATHWAY: Peptidoglycan biosynthesis.
CC -!- SUBCELLULAR LOCATION: Cytoplasmic (By similarity).
CC -!- SIMILARITY: Belongs to the murCDEF family.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL: AE008617; AL030308.1; -
CC PIR: H97269; H97769.
CC HAWAP: MF_08639; atypical; 1.
CC InterPro: IPR000713; Mur_ligase.
CC InterPro: IPR004101; Mur_ligase_C.
CC InterPro: IPR005762; MurD.
CC Pfam: PF01225; Mur_ligase; 1.
CC Pfam: PF02875; Mur_ligase_C; 1.
CC TIGRfams: TIGR01087; murD; 1.
CC Peptidoglycan synthesis; Cell wall; Cell division; Ligase;
KW ATP-binding; Complete proteome.
FT NP BIND 111 117 ATP (POTENTIAL).
FT DOMAIN 260 306 RPE3.
SQ SEQUENCE 500 AA; 55787 MW; 024B268F6074822C CRC64;
Query Match 1.5%; Score 8; DB 1; Length 500;
Best Local Similarity 100.0%; Pred. No. 12;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 85 QGSPAEPE 92
DB 278 QGSPAEPE 285
ID _CALD RAT STANDARD; PRT; 531 AA.
AC Q62736;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 15-MAR-2004 (Rel. 43, Last annotation update)
DE Non-muscle caldesmon (CDW) (I-caldesmon).
GN CALD1.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Rattus.
OX NCBI_TaxID=10116;

```

```

RN [1]
RN SEQUENCE FROM N.A., MUTAGENESIS, AND PHOSPHORYLATION SITES.
RC TISSUE=Liver;
RX MEDLINE=95181370; PubMed=7876150;
RA Yamashiro S., Yamakita Y., Yoshida K.-S., Takiguchi K., Matsumura F.;
RT "Characterization of the COOH terminus of non-muscle caldesmon
RL J. Biol. Chem. 270:4023-4030(1995).
RN [2]
RN PHOSPHORYLATION BY CDC2.
RX MEDLINE=91095023; PubMed=1986309;
RA Yamashiro S., Yamakita Y., Hosoya H., Matsumura F.;
RT "Phosphorylation of non-muscle caldesmon by p34cdc2 kinase during
RL mitosis.";
RL Nature 349:169-172(1991).
CC -!- FUNCTION: Actin- and myosin-binding protein implicated in the
CC regulation of actomyosin interactions in smooth muscle and actin
CC nonmuscle cells (could act as a bridge between myosin and actin
CC filaments). Stimulates actin binding of tropomyosin which
CC increases the stabilization of actin filament structure. In muscle
CC tissues, inhibits the actomyosin ATPase by binding to P-actin.
CC This inhibition is attenuated by calcium-calmodulin and is
CC potentiated by tropomyosin. Interacts with actin, myosin, two
CC molecules of tropomyosin and with calmodulin. Also play an
CC essential role during cellular mitosis and receptor capping.
CC -!- SUBCELLULAR LOCATION: On thin filaments in smooth muscle and on
CC stress fibers in fibroblasts (nonmuscle) (By similarity).
CC -!- TISSUE SPECIFICITY: High-molecular-weight caldesmon (h-caldesmon)
CC is predominantly expressed in smooth muscles, whereas low-
CC molecular-weight caldesmon (l-caldesmon) is widely distributed in
CC non-muscle tissues and cells. Not expressed in skeletal muscle or
CC heart (By similarity).
CC -!- DOMAIN: The N-terminal part seems to be a myosin/calmodulin-
CC binding domain, and the C-terminal a tropomyosin/actin/calmodulin-
CC binding domain. These two domains are separated by a central
CC helical region in the smooth-muscle form.
CC -!- PTM: In non-muscle cells, phosphorylation by CDC2 during mitosis
CC causes caldesmon to dissociate from microfilaments.
CC Phosphorylation reduces caldesmon binding to actin, myosin, and
CC calmodulin as well as its inhibition of actomyosin ATPase
CC activity. Phosphorylation also occurs in both quiescent and
CC dividing smooth muscle cells with similar effects on the
CC interaction with actin and calmodulin and on microfilaments
CC reorganization.
CC -!- SIMILARITY: Belongs to the caldesmon family.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL: U18419; AAA68521.1; -
CC PIR: A55887; A55887.
CC InterPro: IPR006017; Caldesmon.
CC InterPro: IPR006018; Caldesmon_LSP.
CC Pfam: PF02029; Caldesmon; 1.
CC PRINTS: PR01076; CALDESMON.
CC Muscle protein; Actin-binding; Calmodulin-binding; Phosphorylation;
KW Alternative splicing.
FT DOMAIN 20 200 MYOSIN AND CALMODULIN-BINDING (BY
FT SIMILARITY).
FT DOMAIN 303 360 TROPOMYOSIN-BINDING (POTENTIAL).
FT DOMAIN 402 412 TROPOMYOSIN-BINDING (POTENTIAL).
FT DOMAIN 392 424 STRONG ACTIN-BINDING (BY SIMILARITY).
FT DOMAIN 454 460 CALMODULIN-BINDING (BY SIMILARITY).
FT DOMAIN 506 531 WEAK ACTIN-BINDING (BY SIMILARITY).
FT DOMAIN 33 40 POLY-ARG.
FT DOMAIN 180 189 POLY-GLU.
FT DOMAIN 279 282 POLY-ARG.
FT DOMAIN 319 322 POLY-GLU.

```

FT DOMAIN 336 339 POLY-GLU.
FT MOD_RES 249 249 PHOSPHORYLATION (BY CDC2).
FT MOD_RES 462 462 PHOSPHORYLATION (BY CDC2).
FT MOD_RES 468 468 PHOSPHORYLATION (BY CDC2).
FT MOD_RES 491 491 PHOSPHORYLATION (BY CDC2).
FT MOD_RES 497 497 PHOSPHORYLATION (BY CDC2).
FT MOD_RES 527 527 PHOSPHORYLATION (BY CDC2).
FT MUTAGEN 249 249 S->A: DECREASES STRONGLY PHOSPHORYLATION-DEPENDENT ACTIN BINDING.
FT MUTAGEN 462 462 S->A: DECREASES PHOSPHORYLATION-DEPENDENT ACTIN BINDING.
FT MUTAGEN 468 468 T->A: DECREASES PHOSPHORYLATION-DEPENDENT ACTIN BINDING.
FT MUTAGEN 491 491 S->A: DECREASES PHOSPHORYLATION-DEPENDENT ACTIN BINDING.
FT MUTAGEN 497 497 S->A: DECREASES PHOSPHORYLATION-DEPENDENT ACTIN BINDING.
FT MUTAGEN 527 527 S->A: DOES NOT DECREASE PHOSPHORYLATION-DEPENDENT ACTIN BINDING.
FT SEQUENCE 531 AA; 60584 MW; CBCE50271A23829 CRC64;
Query Match 1.5%; Score 8; DB 1; Length 531;
Best Local Similarity 100.0%; Pred. No. 12;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 247 AREEEKR 254
Db 334 AREEEKR 341
RESULT 9
ID_VG50 HSVSA STANDARD; PRT; 535 AA.
AC 001012;
DT 01-APR-1993 (Rel. 25, Created)
DT 01-APR-1993 (Rel. 25, Last sequence update)
DT 01-APR-1993 (Rel. 25, Last annotation update)
DE Probable transcription activator EDRF1.
GN 50 OR HDR21.
OS Herpesvirus saimiri (strain 11).
OC Viruses; dsDNA viruses, no RNA stage; Herpesviridae;
OC Gammaherpesvirinae; Rhadinovirus.
OX NCBI_TaxID=10383;
RN [1]
RP SEQUENCE FROM N.A. PubMed=1321287;
RX MEDLINE=92333688; PubMed=1321287;
RA Albrecht J.-C., Nicholas J., Biller D., Cameron K.R., Biesinger B.,
RA Newman C., Wittmann S., Craxton M.A., Coleman H., Fleckenstein B.,
RA Honess R.N.;
RT "Primary structure of the herpesvirus saimiri genome.";
RL J. Virol. 66:5047-5058(1992).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=9220228; PubMed=1314457;
RA Nicholas J., Coles L.S., Newman C., Honess R.W.;
RT "Regulation of the herpesvirus saimiri (HVS) delayed-early
RT 110-kilodalton promoter by HVS immediate-early gene products and a
RT homolog of the Epstein-Barr virus R trans activator.";
RL J. Virol. 62:2457-2466(1988).
CC -!- FUNCTION: TRANSCRIPTION ACTIVATION. REGULATES THE DELAYED-EARLY
CC 110 kDa PROMOTER.
CC -!- SIMILARITY: TO EBV BRLF1.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its

CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
CC EXBL; X64346; CAA45672.1; ALT INIT.
CC EXBL; M86409; AAA46124.1; ALT INIT.
CC EXBL; M60850; AAA46159.1; ALT INIT.
CC InterPro: IPR004998; Herpes_TAP50.
CC Pfam: PF03326; Herpes_TA50_1.
CC Transcription regulation; Activator; DNA-binding; Early protein.
CC SEQUENCE 535 AA; 60050 MW; B4F2B9ABA38816FB CRC64;
Query Match 1.5%; Score 8; DB 1; Length 535;
Best Local Similarity 100.0%; Pred. No. 12;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 151 GLGKEITL 158
Db 187 GLGKEITL 194
RESULT 10
ID_TFC5 YEAST STANDARD; PRT; 594 AA.
AC P46678;
DT 01-NOV-1995 (Rel. 32, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Transcription factor TFIIB B" component (TFIIB90).
GN TFC5 OR TFC7 OR YNL039W OR N2682.
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycetina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
OX NCBI_TaxID=4932;
RN [1]
RP SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
RC STRAIN=8288C / AB972;
RX MEDLINE=96003864; PubMed=7568218;
RA Kassavetis G.A., Nguyen S.T., Kobayashi R., Kumar A.,
RA Geiduschek E.P., Pisano M.;
RT "Cloning, expression, and function of TFC5, the gene encoding the B'
RT component of the Saccharomyces cerevisiae RNA polymerase III
RT transcription factor TFIIB.";
RL Proc. Natl. Acad. Sci. U.S.A. 92:9786-9790(1995).
RN [2]
RP SEQUENCE FROM N.A.
RX STRAIN=8288C;
RX MEDLINE=96203116; PubMed=8617241;
RA Rueth J., Conesa C., Dieci G., Lefebvre O., Duesterhoeft A.,
RA Otonello S., Sentenac A.;
RT "A suppressor of mutations in the class III transcription system
RT encodes a component of yeast TFIIB.";
RL EMBO J. 15:1941-1949(1996).
RN [3]
RP SEQUENCE FROM N.A.
RX STRAIN=8288C;
RA Roberts S., Miller S., Lane W.S., Hahn S.;
RL Submitted (OCT-1995) to the EMBL/GenBank/DBJ databases.
RN [4]
RP SEQUENCE FROM N.A.
RA Duesterhoeft A., Floeth M., Fritz C., Heuss-Neitzel D., Hilbert H.,
RA Moestl D.;
RL Submitted (MAY-1996) to the EMBL/GenBank/DBJ databases.
CC -!- FUNCTION: General activator of RNA polymerase III transcription.
CC -!- SUBUNIT: TFIIB comprises the TATA-binding protein (TBP), the
CC B-related factor (BRF) and the B" component (TFC5).
CC -!- SUBCELLULAR LOCATION: Nuclear
CC -!- SIMILARITY: TO S.POMBE SPCC1919.14C.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its

use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announcement/> or send an email to license@isb-sib.ch).

EMBL; U31819; AAC49073.1; -;
EMBL; U38415; AAC49364.1; -;
EMBL; U37533; AAC49348.1; -;
EMBL; Z71315; CAR95906.1; -;
PIR; S62141; S62141. -;
GermOnline; 143046; -;
TRANSFAC; T00055; -;
SGD; S0004984; TFC5.
InterPro; IP0001005; Myb DNA-binding.
Pfam; PF00249; myb DNA-binding; 1.
SMART; SM00717; SANT; 1.
KW Transcription regulation; Activator; Nuclear protein.
SQ SEQUENCE 594 AA; 67687 MW; 7B0115BBB2491175 CRC64;

Query Match 1.5%; Score 8; DB 1; Length 594;
Best Local Similarity 100.0%; Pred. No. 13;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 249 EEEKRKE 256
| | | | |
DB 334 EEEKRKE 341

RESULT 11
YF58 YEAST STANDARD; PRT; 756 AA.
AC Q99293;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Hypothetical protein YPL158C.
GN YPL158C OR F2570.
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
OX NCBI_TaxID=4932;
[1]
RN SEQUENCE FROM N.A.
RP STRAIN=S288C / AB972;
RC MEDLINE=97103777; PubMed=8948103;
RX Purnelle B., Coster F., Goffeau A.;
RA "The sequence of 55 kb on the left arm of yeast chromosome XVI
RT identifies a small nuclear RNA, a new putative protein kinase and two
RI new putative regulators.";
RL Yeast 12:1483-1492(1996).
[2]
RN SEQUENCE FROM N.A.
RP STRAIN=S288C / AB972;
RC MEDLINE=9713271; PubMed=9169875;
RX Bussey H., Storms R.K., Ahmed A., Albertmann K., Allen E., Ansoorge W.,
RA Araujo R., Aparicio A., Barrall B.G., Badcock K., Benes V.,
RA Brustin D., Bowman S., Bruckner M., Carpenter J., Cherry J.M.,
RA Chung E., Churcher C.M., Coster F., Davis K., Davis R.W.,
RA Dietrich F.S., Delius H., DiPaolo T., Dubois E., Duesterhoft A.,
RA Duncan M., Floeth M., Fortin N., Friesen J.D., Fritz C., Goffeau A.,
RA Hall J., Hebling U., Heumann K., Hilbert H., Hillier L.W.,
RA Hunicker-Smith S., Hyman R., Johnston M., Kalman S., Kleine K.,
RA Komp C., Kurdi O., Laethkari D., Lew H., Lin A., Lin D., Louis E.J.,
RA Marathe R., Messinguy F., Mees H.-W., Mitiapati S., Moestl D.,
RA Mueller-Auer S., Namath A., Nenwich U., Oefner P., Pearson D.,
RA Petel F.X., Pohl T.M., Purnelle D., Schafer M., Scharfe M.,
RA Scherens B., Schramm S., Schroeder M., Sdicu A.M., Tettelin H.,
RA Urrestarazu L.A., Ushinsky S., Vierendeels F., Vissers S., Voss H.,
RA Walsh S.V., Wambutt R., Wang Y., Wedler E., Wedler E., Winnett E.,
RA Zhong W.W., Zollner A., Vo D.H., Hani J.;
RL "The nucleotide sequence of Saccharomyces cerevisiae chromosome XVI";
RT Nature 387:103-105(1997).

This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announcement/> or send an email to license@isb-sib.ch).

EMBL; X96770; CAA85563.1; -;
EMBL; Z73514; CAA97863.1; -;
PIR; S65169; S65169.
GermOnline; 144140; -;
SGD; S0006079; YPL158C.
KW Hypothetical protein; Coiled coil.
FT DOMAIN 253 263
| | | | |
FT DOMAIN 648 711 COILED COIL (POTENTIAL).
FT DOMAIN 652 658 POLY-GLU.
FT DOMAIN 664 669 POLY-GLU.
FT DOMAIN 686 690 POLY-GLU.
SQ SEQUENCE 758 AA; 84845 MW; 3B4FA92B91C87F2B CRC64;

Query Match 1.5%; Score 8; DB 1; Length 758;
Best Local Similarity 100.0%; Pred. No. 17;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 249 EEEKRKE 256
| | | | |
DB 587 EEEKRKE 594

RESULT 12
RPOB LOTJA STANDARD; PRT; 1070 AA.
AC Q9BB59;
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE DNA-directed RNA polymerase beta chain (EC 2.7.7.6).
GN RPOB.
OS Lotus japonicus.
OC Chloroplast.
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids;
OC eurosids I; Fabales; Fabaceae; Papilionoideae; Loteae; Lotus.
OX NCBI_TaxID=34305;
[1]
RN SEQUENCE FROM N.A.
RP STRAIN=Accession MG-20;
RC MEDLINE=21082923; PubMed=11214967;
RX Kato T., Kaneko T., Sato S., Nakamura Y., Tabata S.;
RA "Complete structure of the chloroplast genome of a legume, Lotus
RT japonicus.";
RL DNA Res. 7:323-330(2000).
CC -1- FUNCTION: DNA-dependent RNA polymerase catalyzes the transcription
CC of DNA into RNA using the four ribonucleoside triphosphates as
CC substrates.
CC -1- CATALYTIC ACTIVITY: N nucleoside triphosphate = N diphosphate +
CC {RNA} (N).
CC -1- SUBUNIT: In chloroplasts, the RNA polymerase is composed of four
CC subunits: alpha, beta, beta', and beta".
CC -1- SUBCELLULAR LOCATION: Chloroplast.
CC -1- SIMILARITY: Belongs to the RNA polymerase beta chain family.
[2]
CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announcement/> or send an email to license@isb-sib.ch).

EMBL; AP002983; BAB3194.1; -;
HSSP; Q9KNU7; 1HQW.

```

DR InterPro; IPR007121; RNA_pol_B.
DR InterPro; IPR007644; RNA_pol_Rpb2_1.
DR InterPro; IPR007642; RNA_pol_Rpb2_2.
DR InterPro; IPR007645; RNA_pol_Rpb2_3.
DR InterPro; IPR007120; RNA_pol_Rpb2_6.
DR InterPro; IPR007641; RNA_pol_Rpb2_7.
DR Pfam; PF04563; RNA_pol_Rpb2_1; 1.
DR Pfam; PF04561; RNA_pol_Rpb2_2; 1.
DR Pfam; PF04565; RNA_pol_Rpb2_3; 1.
DR Pfam; PF00562; RNA_pol_Rpb2_6; 1.
DR Pfam; PF04560; RNA_pol_Rpb2_7; 1.
DR PROSITE; PS01166; RNA_POL_BETA; 1.
KW Transferase; DNA-directed RNA polymerase; Transcription; Chloroplast.
SQ SEQUENCE 1070 AA; 130892 MW; 71268D99B0B5C063 CRC64;

Query Match 1.5%; Score 8; DB 1; Length 1070;
Best Local Similarity 100.0%; Pred.No. 22;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 226 ESLCRELQ 233
Db 241 ESLCRELQ 248

RESULT 13
CARB_THETN
ID CARB_THETN STANDARD; PRT; 1072 AA.
AC Q83EKO;
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Carbamoyl-phosphate synthase large chain (EC 6.3.5.5) (Carbamoyl-
phosphate synthetase ammonia chain).
GN CARB OR TWE0816.
OS Thermoanaerobacter tengcongensis.
OC Bacteria; Firmicutes; Clostridia; Thermoanaerobacteriales;
OC Thermoanaerobacteriaceae; Thermoanaerobacter.
OX NCBI_TaxID=119072;
RN [1]
SEQUENCE FROM N.A.
STRAIN=MB4 / JCM 11007;
MEDLINE=219292816; PubMed=1197336;
BAO Q., Tian Y., Li W., Xu Z., Xuan Z., Hu S., Dong W., Yang J.,
Chen Y., Xu Y., Lai X., Huang L., Dong X., Ma Y., Ling L.,
Tan H., Chen R., Wang J., Yu J., Yang H.;
RA "A complete sequence of T. tengcongensis genome.";
RL Genome Res. 12:689-700(2002).
CC -!- CATALYTIC ACTIVITY: 2 ATP + L-glutamine + CO(2) + H(2)O = 2 ADP +
phosphate + L-glutamate + carbamoyl phosphate.
CC -!- COFACTOR: Binds 3 manganese ions per subunit (By similarity).
CC -!- PATHWAY: Arginine biosynthesis.
CC -!- SUBUNIT: Composed of two chains; the small (or glutamine) chain
promotes the hydrolysis of glutamine to ammonia, which is used by
the large (or ammonia) chain to synthesize carbamoyl phosphate (By
similarity).
CC -!- SIMILARITY: Belongs to the carb family.

This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
EMBL; AE013048; AM24073.1; -.
DR HAVAP; MF_01210; -.
DR InterPro; IPR006275; CarA_L_glu.
DR InterPro; IPR005483; CPhase_L.
DR InterPro; IPR005479; CPhase_L_D2.
DR InterPro; IPR005480; CPhase_L_D3.
DR InterPro; IPR005481; CPhase_L_N.

```

```

DR InterPro; IPR004362; MGS-like.
DR Pfam; PF00289; CPhase_L_Chain; 2.
DR Pfam; PF02786; CPhase_L_D2; 2.
DR Pfam; PF02787; CPhase_L_D3; 1.
DR Pfam; PF02142; MGS; 1.
DR PRINTS; PR00098; CPhase.
DR TIGRfam; TIGR01369; CPhaseII_lrg; 1.
DR PROSITE; PS00866; CPhase_1; 1.
DR PROSITE; PS00867; CPhase_2; 2.
KW Arginine biosynthesis; Pyrimidine biosynthesis; Ligase; Repeat;
KW ATP-binding; Manganese; Complete proteome.
FT DOMAIN 1 401
FT DOMAIN 402 546
FT DOMAIN 547 929
FT DOMAIN 930 1072
FT REPEAT 547 1072
FT NP_BIND 153 210
FT NP_BIND 302 352
FT METAL 284 284
FT METAL 298 298
FT METAL 300 300
FT METAL 820 820
FT METAL 832 832
SQ SEQUENCE 1072 AA; 119000 MW; D5F08AD6C8BE75D37 CRC64;

Query Match 1.5%; Score 8; DB 1; Length 1072;
Best Local Similarity 100.0%; Pred.No. 22;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 238 SLKEGVQ 245
Db 36 SLKEGVQ 43

RESULT 14
TRAA_AGRTS
ID TRAA_AGRTS STANDARD; PRT; 1100 AA.
AC Q44349;
DT 01-NOV-1997 (Rel. 35, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Conjugal transfer protein traa.
GN TRAA OR ATU6127 OR AGR_PTI_237.
OS Agrobacterium tumefaciens (strain C58 / ATCC 33970).
OC Plasmid pTiC58.
OC Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;
OC Rhizobiaceae; Rhizobium/Agrobacterium group; Agrobacterium.
OX NCBI_TaxID=176299;
RN [1]
SEQUENCE FROM N.A.
MEDLINE=96312367; PubMed=8763953;
RA Farrand S.K., Hwang I., Cook D.M.;
RA "The tra region of the nopaline-type Ti plasmid is a chimera with
elements related to the transfer systems of RSP1010, RP4, and F.";
RL J. Bacteriol. 178:4233-4247(1996).
RN [2]
SEQUENCE FROM N.A.
MEDLINE=21608550; PubMed=11743193;
RA Wood D.W., Setubal J.C., Kaul R., Monks D.E., Kitajima J.P.,
Wood V.K., Zhou Y., Chen L., Wood G.B., Almeida N.P. Jr., Woo L.,
Chen Y., Paulsen I.T., Eisen J.A., Karp P.D., Bovee D. Sr.,
Chapman P., Clendenning J., Deatherage G., Gillet W., Grant C.,
Kutyavin T., Levy R., Li M.-J., McClelland E., Palmieri A., Gordon D.,
Raymond C., Rouse G., Saenphimachak C., Wu Z., Romero P., Krespan W., Perry M.,
Zhang S., Yoo H., Biddle P., Jung M., Hendrick C., Zhao Z.-Y., Doan M.,
Chumley F., Tingey S.V., Tomb J.-F., Gordon M.P., Olson M.V.,
Nester E.W.;
RA "The genome of the natural genetic engineer Agrobacterium tumefaciens
C58.";
RL Science 294:2317-2323(2001).
RN [3]

```

```
RP SEQUENCE FROM N.A.
RX MEDLINE=21608551; PubMed=11743194;
RA Goodner B., Hinkle G., Gattung S., Miller N., Blanchard M.,
RA Ouello B., Goldman B.S., Cao Y., Askenazi M., Halling C., Mullin L.,
RA Houmlet K., Gordon J., Vaudin M., Iartchouk O., Epp A., Liu F.,
RA William C., Allinger W., Doughty D., Scott C., Lippas C., Markez B.,
RA Flanagan C., Crowell C., Gursen J., Lomo C., Sear C., Strub G.,
RA Cielo C., Slater S.,
RT "Genome sequence of the plant pathogen and biotechnology agent
RT Agrobacterium tumefaciens C58."
RL Science 294:2323-2328(2001).
CC 1- SIMILARITY: Belongs to the mobA/mobL family.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC -----
DR EMBL; AF010180; AAC17212.1; -.
DR EMBL; AF009431; AAL46363.1; -.
DR EMBL; AF007939; AAX91091.1; ALT_INIT.
DR PIR; AE3243; AE3243.
DR PIR; T03419; T03419.
DR InterPro; IPR005053; MobA_MobL.
DR Pfam; PF03389; MobA_MobL; 1.
KW Conjugation; ATP-binding; Plasmid; Complete proteome.
FT NP_BIND 404 411
FT CONFLICT 372 372 L -> LL (IN REF. 1).
SQ SEQUENCE 1100 AA; 123477 MW; B85D31DB526B7344 CRC64;

Query Match 1.5%; Score 8; DB 1; Length 1100;
Best Local Similarity 100.0%; Pred. No. 23;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 135 VGDHRHR 142
DB 671 VGDHRHR 678
|||||

RESULT 15
MY10 BOVIN STANDARD; PRT; 2052 AA.
ID MY10 BOVIN
AC P79114;
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Myosin X.
GN MYO10.
OS Bos taurus (Bovine).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
OC Bovidae; Bovinae; Bos.
OX NCBI_TaxID=9913;
RN [1]
RP SEQUENCE FROM N.A.
RX TISSUE=Aorta;
RC MEDLINE=20442246; PubMed=10984435;
RA Berg J.S., Derfler B.H., Pennisi C.M., Corey D.P., Cheney R.B.;
RT "Myosin-X, a novel myosin with pleckstrin homology domains, associates
RT with regions of dynamic actin."
RL J. Cell Sci. 113:3439-3451(2000).
CC 1- FUNCTION: Myosins are actin-based motor molecules with ATPase
CC activity. Unconventional myosins serve in intracellular movements.
CC Their highly divergent tails are presumed to bind to membranous
CC compartments, which would be moved relative to actin filaments (By
CC similarity). Plays a role in regions of dynamic actin.
CC 1- SIMILARITY: Contains 1 myosin-like globular head domain.
CC 1- SIMILARITY: Contains 1 FERM domain.
CC 1- SIMILARITY: Contains 3 IQ domains.
CC 1- SIMILARITY: Contains 1 MYTH4 domain.

CC 1- SIMILARITY: Contains 2 PH domains.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC -----
DR EMBL; U55042; AAB39486.1; -.
DR PIR; T18519; T18519.
DR HSPF; P08799; LMND.
DR InterPro; IPR000299; Band_4.1.
DR InterPro; IPR000048; IQ_region.
DR InterPro; IPR001609; myosin_head.
DR InterPro; IPR000857; MYTH4.
DR InterPro; IPR001849; PH.
DR Pfam; PF00612; IQ; 3.
DR Pfam; PF00063; myosin_head; 1.
DR Pfam; PF00784; MYTH4; 1.
DR Pfam; PF00169; PH; 3.
DR PRINTS; PR00193; MYOSINHEAVY.
DR ProDom; PD000355; myosin_head; 1.
DR SMART; SM00295; B41; 1.
DR SMART; SM00015; IQ; 3.
DR SMART; SM00242; MISC; 1.
DR SMART; SM00139; MYTH4; 1.
DR SMART; SM00233; PH; 2.
DR PROSITE; PS00660; FERM_1; FALSE_NEG.
DR PROSITE; PS00661; FERM_2; FALSE_NEG.
DR PROSITE; PS00057; FERM_3; 1.
DR PROSITE; PS50096; IQ; 2.
DR PROSITE; PS50003; PH_DOMAIN; 2.
KW Myosin; ATP-binding; Actin-binding; Coiled coil; Repeat.
FT DOMAIN 1 727 MYOSIN HEAD-LIKE.
FT DOMAIN 742 763 IQ 1.
FT DOMAIN 764 787 IQ 2.
FT DOMAIN 788 817 IQ 3.
FT DOMAIN 1206 1304 PH 1.
FT DOMAIN 1386 1491 PH 2.
FT DOMAIN 1694 2038 FERM.
FT DOMAIN 800 941 COILED COIL (POTENTIAL).
SQ SEQUENCE 2052 AA; 235837 MW; 43DF13424B4B2D28 CRC64;

Query Match 1.5%; Score 8; DB 1; Length 2052;
Best Local Similarity 100.0%; Pred. No. 40;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 248 REEEKRK 255
DB 824 REEEKRK 831
|||||

RESULT 16
RK21 ODOSI STANDARD; PRT; 105 AA.
ID RK21 ODOSI
AC P49557;
DT 01-FEB-1996 (Rel. 33, Created)
DT 01-FEB-1996 (Rel. 33, Last sequence update)
DT 01-FEB-1996 (Rel. 33, Last annotation update)
DE Chloroplast 50S ribosomal protein L21.
GN RPL21.
OS Odontella sinensis (Marine centric diatom).
OC Chloroplast.
OC Eukaryota; Stramenopiles; Bacillariophyta; Coscinodiscophyceae;
OC Biddulphiophycidae; Eupodiscaceae; Odontella.
OX NCBI_TaxID=2839;
RN [1]
RP SEQUENCE FROM N.A.
RA Kowalik K.V., Stoebe B., Schaffran I., Kroth-Pancic P., Freier U.;
RT "The chloroplast genome of a chlorophyll a+c-containing alga,
RT Odontella sinensis.";
```



```

RL Plant Mol. Biol. Rep. 13:336-342(1995).
CC -!- FUNCTION: This protein binds to 23S ribosomal RNA in the presence
CC of protein L20 (by similarity).
CC -!- SIMILARITY: Belongs to the L21P family of ribosomal proteins.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; Z67753; CAA91663.1; -
CC PIR; S78290; S78290.
CC InterPro; IPR001787; Ribosomal_L21P.
CC Pfam; PF00829; Ribosomal_L21P; 1.
CC ProDom; PD003604; Ribosomal_L21P; 1.
CC TIGRfam; TIGR00061; L21; 1.
CC PROSITE; PS01169; RIBOSOMAL_L21; 1.
CC Ribosomal protein; rRNA-binding; Chloroplast.
CC SEQUENCE 105 AA; 12428 MW; ED99D758B9BCD008 CRC64;
CC -----
CC Query Match 1.3%; Score 7; DB 1; Length 105;
CC Best Local Similarity 100.0%; Pred. No. 30;
CC Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
CC -----
CC QY 152 LGKEITL 158
CC |||||
CC Db 30 LGKEITL 36
CC -----
RESULT 17
INSC_ECOLI
ID INSC_ECOLI STANDARD; PRT; 121 AA.
AC P19776; O07989; O08018; O08019; P76357; P77346;
DT 01-FEB-1991 (Rel. 17, Created)
DT 01-FEB-1991 (Rel. 17, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Transposase insc for insertion element IS2A/D/F/H/I/K.
GN (INSC1 OR B0360) AND (INSC2 OR B1403) AND (INSC3 OR B1997) AND
GN (INSC4 OR B2861) AND (INSC5 OR B3044) AND (INSC6 OR B4272).
OS Escherichia coli.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Escherichia.
OX NCBI_TaxID=562;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=K12;
RX MEDLINE=88137965; PubMed=2830172;
RA Ronsker H.J., Rak B.;
RT "Genetic organization of insertion element IS2 based on a revised
RT nucleotide sequence.";
RL Gene 59:291-296(1987).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=K12 / MG1655;
RX MEDLINE=95334362; PubMed=7610040;
RA Burland V.D., Plunkett G. III, Sofia H.J., Daniels D.L.,
RA Blattner F.R.;
RT "Analysis of the Escherichia coli genome VI: DNA sequence of the
RT region from 92.8 through 100 minutes.";
RL Nucleic Acids Res. 23:2105-2119(1995).
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=K12 / MG1655;
RX MEDLINE=97426617; PubMed=9278503;
RA Blattner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,
RA Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.P.,
RA Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,
RA Mau B., Shao Y.;
RT "The complete genome sequence of Escherichia coli K-12.";
RT Science 277:1453-1474(1997).

```

```

RN SEQUENCE FROM N.A.
RC STRAIN=K12;
RX MEDLINE=97251357; PubMed=9097039;
RA Aiba H., Baba T., Fujita K., Hayashi K., Inada T., Isono K.,
RA Itoh T., Kasai H., Kashimoto K., Kimura S., Kitakawa M., Mori T.,
RA Kitagawa M., Makino K., Miki T., Mizobuchi K., Mori H., Mori T.,
RA Motomura K., Nakade S., Nakamura Y., Nishimoto H., Nishio Y.,
RA Oshima T., Saito N., Sampel G., Seki Y., Sivasubram S.,
RA Tagami H., Takeda J., Takemoto K., Takeuchi Y., Wada C.,
RA Yamamoto Y., Horiuchi T.;
RT "A 570-kb DNA sequence of the Escherichia coli K-12 genome
RT corresponding to the 28.0-40.1 min region on the linkage map.";
RL DNA Res. 3:363-377(1996).
RN [5]
RP SEQUENCE FROM N.A.
RC STRAIN=K12;
RX MEDLINE=97251358; PubMed=9097040;
RA Itoh T., Aiba H., Baba T., Fujita K., Hayashi K., Inada T., Isono K.,
RA Kasai H., Kimura S., Kitakawa M., Kitagawa M., Makino K., Miki T.,
RA Mizobuchi K., Mori H., Mori T., Motomura K., Nakade S., Nakamura Y.,
RA Nishimoto H., Nishio Y., Oshima T., Saito N., Sampel G., Seki Y.,
RA Sivasubram S., Tagami H., Takeda J., Takemoto K., Wada C.,
RA Yamamoto Y., Horiuchi T.;
RT "A 460-kb DNA sequence of the Escherichia coli K-12 genome
RT corresponding to the 40.1-50.0 min region on the linkage map.";
RL DNA Res. 3:379-392(1996).
CC -!- FUNCTION: Involved in the transposition of the insertion sequence
CC IS2.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; V00279; CAA23542.1; -
CC EMBL; U14003; AAA97168.1; ALT INIT.
CC EMBL; U28377; AAA69212.1; ALT INIT.
CC EMBL; U28375; AAA83043.1; ALT INIT.
CC EMBL; AE000143; AAC73463.1; ALT INIT.
CC EMBL; AE000237; AAC74485.1; ALT INIT.
CC EMBL; AE000291; AAC75058.1; ALT INIT.
CC EMBL; AE000369; AAC75900.1; -
CC EMBL; AE000386; AAC76080.1; -
CC EMBL; AE000498; AAC77228.1; ALT INIT.
CC EMBL; D90778; BAA15013.1; ALT INIT.
CC EMBL; D90779; BAA15019.1; ALT INIT.
CC EMBL; D90838; BAA15822.1; ALT INIT.
CC EMBL; D90850; BAA16005.1; ALT INIT.
CC EMBL; D90851; BAA16013.1; ALT INIT.
CC EMBL; D90852; BAA16036.1; ALT INIT.
CC EcoGene; EG40003; insc.
CC InterPro; IPR002514; Transposase_8.
CC Pfam; PF01527; Transposase_8; 1.
CC KW Transposable element; Transposition; DNA-binding; DNA recombination;
CC Complete proteome.
CC VARIANT 34 34 L -> F (IN B1997).
CC SEQUENCE 121 AA; 13452 MW; 59431E5C452E067A CRC64;
CC -----
CC Query Match 1.3%; Score 7; DB 1; Length 121;
CC Best Local Similarity 100.0%; Pred. No. 34;
CC Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
CC -----
CC QY * 346 LLKEAVE 352
CC |||||
CC Db 96 LLKEAVE 102
CC -----
RESULT 18
INSC_SHIFL

```

```
Query Match      1.3%; Score 7; DB 1; Length 121;
Best Local Similarity 100.0%; Pred. No. 34;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
FT NON_TER 1 1
FT SIGNAL <1 7
FT CHAIN 8 127
FT ACT_SITE 18 18
FT SITE 62 66
FT DISULFID 80 90
FT DISULFID 104 124
FT CONFLICT 25 25
SQ SEQUENCE 127 AA; 14039 MW; 78F70158B7525853 CRC64;

Query Match 1.3%; Score 7; DB 1; Length 127;
Best Local Similarity 100.0%; Pred. No. 35;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 241 BEGVORA 247
DB 27 BEGVORA 33

RESULT 20
NL13 PARJU
ID NL13 PARJU STANDARD; PRT; 138 AA.
AC Q40305;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Probable nonspecific lipid-transfer protein 1 precursor (LTP) (Major
DE pollen allergen Par j 1.0201) (Par j 1) (P1 protein).
OS Parietaria judaica.
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids;
OC eurosids I; Rosales; Urticaceae; Parietaria.
OX NCBI_TaxID=33127;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Pollen.
RX MEDLINE=97258596; PubMed=9104790;
RA Duro G., Colombo P., Costa M.A., Izso V., Porcasi R., di Fiore R.,
RA Locorotondo G., Cocchiara R., Geraci D.;
RT "Isolation and characterization of two cDNA clones coding for
RT isoforms of the Parietaria judaica major allergen Par j 1.0101.";
RL Int. Arch. Allergy Immunol. 112:348-355(1997).
CC -!- FUNCTION: Plant nonspecific lipid-transfer proteins transfer
CC phospholipids as well as galactolipids across membranes. May play
CC a role in wax or cutin deposition in the cell walls of expanding
CC epidermal cells and certain secretory tissues.
CC -!- ALLERGEN: Causes an allergic reaction in human. Can bind IgE and
CC induce histamine release from basophils of Pj-pollen-allergic
CC subjects.
CC -!- SIMILARITY: Belongs to the plant LTP family.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
```

SQ SEQUENCE 139 AA; 14986 MW; 36C3544DD699D6C7 CRC64;
Query Match 1.3%; Score 7; DB 1; Length 139;
Best Local Similarity 100.0%; Pred. No. 38;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 116 EKEPSKG 122
DB 22 EKEPSKG 28
|||||

RESULT 22
ID_NL12_PARJU STANDARD; PRT; 176 AA.
AC 004404;
DT 15-DEC-1998 (Rel. 37, Created)
DT 15-DEC-1998 (Rel. 37, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Probable nonspecific lipid-transfer protein 1 precursor (LTP) (Major
DE pollen allergen par j 1.0102) (Par j 1) (99 protein).
OS Parietaria judaica.
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosids;
OC eurosids I; Rosales; Urticales; Parietaria.
OX NCBI_TaxID=33127;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Pollen;
RX MEDLINE=97258596; PubMed=9104790;
RA Duro G., Colombo P., Costa M.A., Izzo V., Porcasi R., di Fiore R.,
RA Locorotondo G., Cocchiara R., Geraci D.;
RT "Isolation and characterization of two cDNA clones coding for
RT isoforms of the Parietaria judaica major allergen Par j 1.0101.";
RL Int. Arch. Allergy Immunol. 112:348-355(1997).
CC -1- FUNCTION: Plant nonspecific lipid-transfer proteins transfer
CC phospholipids as well as galactolipids across membranes. May play
CC a role in wax or cutin deposition in the cell walls of expanding
CC epidermal cells and certain secretory tissues.
CC -1- ALLERGEN: Causes an allergic reaction in human. Can bind IgE and
CC induce histamine release from basophils of Pj-pollen-allergic
CC subjects.
CC -1- SIMILARITY: Belongs to the plant LTP family.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).

DR EMBL; X95867; CAA65123.1;
DR HSP; P19656; IMZM.
DR InterPro; IPR003612; AAI.
DR Pfam; PF00234; tlyp_alpha_aml1; 1.
DR PRINTS; PR00382; LIPIDTRANSFER.
DR SMART; SM00499; AAI; 1.
DR PROSITE; PS00597; PLANT_LTP; FALSE_NEG.
KW Lipid-binding; Transport; Multigene family; Signal; Allergen.
FT SIGNAL 1 37
FT CHAIN 38 176
FT PROBABLE NONSPECIFIC LIPID-TRANSFER
FT PROTEIN 1.
FT DISULFID 41 89 BY SIMILARITY.
FT DISULFID 51 66 BY SIMILARITY.
FT DISULFID 67 112 BY SIMILARITY.
FT DISULFID 87 128 BY SIMILARITY.
SQ SEQUENCE 176 AA; 18454 MW; 3AC31635036AE323 CRC64;

Query Match 1.3%; Score 7; DB 1; Length 176;
Best Local Similarity 100.0%; Pred. No. 46;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 116 EKEPSKG 122
DB 59 EKEPSKG 65
|||||

RESULT 23
ID_RRF_WOLSU STANDARD; PRT; 186 AA.
AC 07MAD8;
DT 15-MAR-2004 (Rel. 43, Created)
DT 15-MAR-2004 (Rel. 43, Last sequence update)
DT 15-MAR-2004 (Rel. 43, Last annotation update)
DE Ribosome recycling factor (Ribosome releasing factor) (RRF).
GN RRF OR WS0314.
OS Wolinella succinogenes.
OC Bacteria; Proteobacteria; Epsilonproteobacteria; Campylobacteriales;
OC Helicobacteraceae; Wolinella.
OX NCBI_TaxID=844;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=DSMZ 1740;
RX MEDLINE=22882897; PubMed=14500908;
RA Baar C., Espinger M., Raddatz G., Simon J., Lanz C., Klimmek O.,
RA Mandakumar R., Gross R., Rosinus A., Keller H., Jagtap P., Linke B.,
RA Meyer F., Lederer H., Schuster S.C.;
RT "Complete genome sequence and analysis of Wolinella succinogenes.";
RL Proc. Natl. Acad. Sci. U.S.A. 100:11690-11695(2003).
CC -1- FUNCTION: Responsible for the release of ribosomes from messenger
CC RNA at the termination of protein biosynthesis. May increase the
CC efficiency of translation by recycling ribosomes from one round of
CC translation to another (By similarity).
CC -1- SUBCELLULAR LOCATION: Cytoplasmic (By similarity).
CC -1- SIMILARITY: Belongs to the RRF family.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).

DR EMBL; BX571657; CAE09465.1;
DR HAMAP; MF 00040; -1.
KW Protein biosynthesis; Complete proteome.
SQ SEQUENCE 186 AA; 21114 MW; E1FAD7F28D09392 CRC64;

Query Match 1.3%; Score 7; DB 1; Length 186;
Best Local Similarity 100.0%; Pred. No. 49;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 405 KIKKLEK 411
DB 139 KIKKLEK 145
|||||

RESULT 24
ID_ISPZ_VIBPA STANDARD; PRT; 187 AA.
AC Q87NA5;
DT 10-OCT-2003 (Rel. 42, Created)
DT 10-OCT-2003 (Rel. 42, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Probable intracellular septation protein.
GN ISPZ OR VP1970.
OS Vibrio parahaemolyticus.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Vibrionales;
OC Vibrionaceae; Vibrrio.
OX NCBI_TaxID=670;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=RIMD 2210633 / Serotype O3:K6;
RX MEDLINE=22508454; PubMed=12620739;

```

RA Makino K., Oshima K., Kurokawa K., Yokoyama K., Uda T., Tagomori K.,
RA Iijima Y., Najima M., Nakano M., Yamashita A., Kubota Y., Kimura S.,
RA Yasunaga T., Honda T., Shinagawa H., Hattori M., Iida T.;
RT "Genome sequence of Vibrio parahaemolyticus: a pathogenic mechanism
RT distinct from that of V. cholerae.";
RL Lancet 361:743-749(2003).
CC -!- FUNCTION: Involved in cell division; probably involved in
CC intracellular septation (By similarity).
CC -!- SUBCELLULAR LOCATION: Integral membrane protein (By similarity).
CC -!- SIMILARITY: Belongs to the ispZ family.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AP005079; BAC60233.1; -.
DR HAMAP; MF 00189; -.
DR InterPro; IPR006008; SeptationA.
DR Pfam; PF04279; IspA; 1.
KW Cell division; Septation; Transmembrane; Complete proteome.
FT TRANSMEM 22 44 Potential.
FT TRANSMEM 51 68 Potential.
FT TRANSMEM 78 100 Potential.
FT TRANSMEM 113 135 Potential.
FT TRANSMEM 150 169 Potential.
SQ SEQUENCE 187 AA; 21317 MW; 23AF8FF42DCBD025 CRC64;

Query Match 1.3%; Score 7; DB 1; Length 187;
Best Local Similarity 100.0%; Pred. No. 49;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 152 LGKEITL 158
DB 104 LGKEITL 110
|||||
RESULT 25
ISPZ_VIBVU STANDARD; PRT; 187 AA.
AC PS9366;
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Probable intracellular septation protein.
GN ISPZ OR VJ3073.
OS Vibrio vulnificus.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Vibrionales;
OC Vibrionaceae; Vibrio.
OX NCBI_TaxID=672;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CMCP6;
RA Rhee J.H., Kim S.Y., Chung S.S., Kim J.J., Moon Y.H., Jeong H.,
RA Chey H.E.;
RT "Complete genome sequence of Vibrio vulnificus CMCP6.";
RL Submitted (DEC-2002) to the EMBL/GenBank/DBJ databases.
CC -!- FUNCTION: Involved in cell division; probably involved in
CC intracellular septation (By similarity).
CC -!- SUBCELLULAR LOCATION: Integral membrane protein (By similarity).
CC -!- SIMILARITY: Belongs to the ispZ family.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----

```

```

DR EMBL; AE016807; AAO11397.1; ALT_INIT.
DR HAMAP; MF 00189; -.
DR InterPro; IPR006008; SeptationA.
DR Pfam; PF04279; IspA; 1.
DR ProDom; PD016710; SeptationA; 1.
DR TIGRPFAM; TIGR00997; ispZ; 1.
KW Cell division; Septation; Transmembrane; Complete proteome.
FT TRANSMEM 22 44 Potential.
FT TRANSMEM 51 68 Potential.
FT TRANSMEM 78 100 Potential.
FT TRANSMEM 113 135 Potential.
FT TRANSMEM 150 169 Potential.
SQ SEQUENCE 187 AA; 21014 MW; AB803DB2398A9FC CRC64;

Query Match 1.3%; Score 7; DB 1; Length 187;
Best Local Similarity 100.0%; Pred. No. 49;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 152 LGKEITL 158
DB 104 LGKEITL 110
|||||
RESULT 26
LOLB_NEIMA STANDARD; PRT; 193 AA.
ID _LOLB_NEIMA
AC P57023;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Outer-membrane lipoprotein lolB precursor.
GN LOLB OR NMA1091.
OS Neisseria meningitidis (serogroup A).
OC Bacteria; Proteobacteria; Betaproteobacteria; Neisseriales;
OC Neisseriaceae; Neisseria.
OX NCBI_TaxID=65699;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=22491 / Serogroup A / Serotype 4A;
RX MEDLINE=20222556; PubMed=10761919;
RA Parkhill J., Achtman M., James K.D., Bentley S.D., Churcher C.,
RA Klee S.R., Morelli G., Basham D., Brown D., Chillingworth T.,
RA Davies R.M., Davis P., Devlin K., Feltwell T., Hamlin N., Holroyd S.,
RA Jagels K., Leather S., Moule S., Mungall K., Quail M.A.,
RA Rajadream M.A., Rutherford K.M., Simmonds M., Skelton J.,
RA Whitehead S., Spratt B.G., Barrell B.G.;
RT "Complete DNA sequence of a serogroup A strain of Neisseria
RT meningitidis 22491.";
RL Nature 404:502-506(2000)
CC -!- FUNCTION: Plays a critical role in the incorporation of
CC lipoproteins in the outer membrane after they are released by the
CC lola protein (By similarity).
CC -!- SUBCELLULAR LOCATION: Attached to the outer membrane by a lipid
CC anchor (By similarity).
CC -!- SIMILARITY: Belongs to the lolB family.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AL162755; CAB84354.1; -.
DR PIR; H81874; H81874.
DR HAMAP; MF 00233; -.
DR InterPro; IPR004565; LolB.
DR ProDom; IPR000437; Prok_lipoprot_S.
DR Pfam; PF03550; LolB; 1.
DR PROSITE; PS00013; PROKAR_LIPOPROTEIN; 1.
KW Chaperone; Outer membrane; Lipoprotein; Transport; Protein transport;
KW Signal; Complete proteome; Palmitate.

```

FT SIGNAL 1 15 BY SIMILARITY.
FT CHAIN 16 193 OUTER-MEMBRANE LIPOPROTEIN LOLB.
FT LIPID 16 16 S-diacylglycerol cysteine (By
similarity).
FT LIPID 16 16 N-palmitoyl cysteine (By similarity).
SQ SEQUENCE 193 AA; 21156 MW; 7259427B44B4P478 CRC64;
Query Match 1.3%; Score 7; DB 1; Length 193;
Best Local Similarity 100.0%; Pred. No. 50;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 57 EELSROL 63
DB 102 EELSROL 108
|||||

RESULT 27
LOLB NEIMB STANDARD; PRT; 193 AA.
AC P57024;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Outer-membrane lipoprotein loloB precursor.
GN LOLB OR NMB0873.
OS Neisseria meningitidis (serogroup B).
OC Bacteria; Proteobacteria; Betaproteobacteria; Neisseriales;
OC Neisseriaceae; Neisseria.
OX NCBI_TaxID=491;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=MC58 / Serogroup B;
RX MEDLINE=20175755; PubMed=10710307;
RA Tettelin H., Saunders N.J., Heidelberg J., Jeffries A.C., Nelson K.E.,
RA Eisen J.A., Karchum K.A., Hood D.W., Peden J.F., Dodson R.J.,
RA Nelson W.C., Gwinn M.L., DeBoy R., Peterson J.D., Hickey E.K.,
RA Haft D.H., Salzberg S.L., White O., Fleischmann R.D., Dougherty B.A.,
RA Mason T., Clecko A., Parksey D.S., Blair E., Cifton H., Clark E.B.,
RA Cotton M.D., Uterback T.R., Khouri H., Qin H., Vamathevan J.,
RA Gill J., Scarlato V., Masignani V., Pizsa M., Grandi G., Sun L.,
RA Smith H.O., Fraser C.M., Moxon E.R., Rappuoli R., Venter J.C.,
RT "Complete genome sequence of Neisseria meningitidis serogroup B strain
MC58.";
RL Science 287:1809-1815(2000).
CC -!- FUNCTION: Plays a critical role in the incorporation of
lipoproteins in the outer membrane after they are released by the
lola protein (By similarity).
CC -!- SUBCELLULAR LOCATION: Attached to the outer membrane by a lipid
anchor (By similarity).
CC -!- SIMILARITY: Belongs to the loloB family.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>)
or send an email to license@isb-sib.ch.

CC EMBL; AE002439; AAP41284.1; -.
DR PIR; A81149; A81149.
DR TIGR; NMB0873; -.
DR HAMAP; MF_00233; -; 1.
DR InterPro; IPR004565; LOLB.
DR InterPro; IPR000437; Prok_lipoprot_S.
DR Pfam; PF03550; LOLB; 1.
DR PROSITE; PS00013; PROKAR_LIPOPROTEIN; 1.
KW Chapterone; Outer membrane; lipoprotein; Transport; Protein transport;
Signal; Complete proteome; Palmitate.
FT SIGNAL 1 15 BY SIMILARITY.
FT CHAIN 16 193 OUTER-MEMBRANE LIPOPROTEIN LOLB.
FT LIPID 16 16 S-diacylglycerol cysteine (By
similarity).

FT SIGNAL 1 16 N-palmitoyl cysteine (By similarity).
SQ SEQUENCE 193 AA; 21141 MW; 4C2E97734BB2C41E CRC64;
Query Match 1.3%; Score 7; DB 1; Length 193;
Best Local Similarity 100.0%; Pred. No. 50;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 57 EELSROL 63
DB 102 EELSROL 108
|||||

RESULT 28
AANT_HDV3 STANDARD; PRT; 195 AA.
AC P29996;
DT 01-APR-1993 (Rel. 25, Created)
DT 01-APR-1993 (Rel. 25, Last sequence update)
DT 01-JUL-1993 (Rel. 26, Last annotation update)
DE Delta antigen.
OS Hepatitis delta virus (isolate D380) (HDV).
OC Viruses; Deltavirus.
OX NCBI_TaxID=31762;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=90324949; PubMed=2374010;
RA Saldanha J.A., Thomas H.C., Monjardino J.P.,
RT "Cloning and sequencing of RNA of hepatitis delta virus isolated from
human serum.";
RL J. Gen. Virol. 71:1603-1606(1990).
CC -!- PTM: Phosphorylated.
CC -!- SUBCELLULAR LOCATION: Nuclear.
CC -!- DISEASE: BINDS SPECIFICALLY TO ANTIGEN FROM PATIENTS WITH CHRONIC
HEPATITIS DELTA VIRAL INFECTIONS.
CC -!- MISCELLANEOUS: Delta antigen binds specifically to HDV RNA.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>)
or send an email to license@isb-sib.ch.

CC EMBL; D01075; BAA00874.1; -.
DR PIR; A35219; SAVLH1.
DR InterPro; IPR002506; HDV ag.
DR Pfam; PF01517; HDV ag; 1.
DR ProDom; PD002887; HDV ag; 1.
KW Antigen; RNA-binding; Nuclear protein; Phosphorylation.
SQ SEQUENCE 195 AA; 21937 MW; 8F09A9C62E9D2721 CRC64;

QY 404 KKIKKLE 410
DB 39 KKIKKLE 45
|||||

RESULT 29
AANT_HDV1 STANDARD; PRT; 195 AA.
AC P29833;
DT 01-APR-1993 (Rel. 25, Created)
DT 01-APR-1993 (Rel. 25, Last sequence update)
DT 01-APR-1993 (Rel. 25, Last annotation update)
DE Delta antigen.
OS Hepatitis delta virus (isolate Lebanon-1) (HDV).
OC Viruses; Deltavirus.
OX NCBI_TaxID=31763;
RN [1]

```
RP SEQUENCE FROM N.A.
RX MEDLINE=92230225; PubMed=1566577;
RA Lee C.M., Bih P.Y., Chao Y.C., Govindarajan S., Lai M.M.C.;
RT "Evolution of hepatitis delta virus RNA during chronic infection.";
RL Virology 188:265-273(1992).
CC -!- SUBCELLULAR LOCATION: Nuclear.
CC -!- PTM: Phosphorylated.
CC -!- DISEASE: BINDS SPECIFICALLY TO ANTISERA FROM PATIENTS WITH CHRONIC
CC HEPATITIS DELTA VIRAL INFECTIONS.
CC -!- MISCELLANEOUS: Delta antigen binds specifically to HDV RNA.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; M84917; -; NOT_ANNOTATED_CDS.
CC F01; A40247; SAVILL1.
CC InterPro; IPR002506; HDV_ag.
CC Pfam; PF01517; HDV_ag; 1.
CC ProDom; PD002887; HDV_ag; 1.
KW Antigen; RNA-binding; Nuclear protein; Phosphorylation.
SQ SEQUENCE 195 AA; 21846 MW; 2B8929C595B5D41B CRC64;
Query Match 1.3%; Score 7; DB 1; Length 195;
Best Local Similarity 100.0%; Pred. No. 51;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 404 KKIKKLE 410
DB 39 KKIKKLE 45
RESULT 30
AANT HDVM1
ID AANT HDVM1 STANDARD; PRT; 195 AA.
AC P25881;
DT 01-MAY-1992 (Rel. 22, Created)
DT 01-MAY-1992 (Rel. 22, Last sequence update)
DT 01-MAY-1992 (Rel. 22, Last annotation update)
DE Delta antigen.
OS Hepatitis delta virus (isolate Japanese M-1) (HDV).
OC Viruses; Deltavirus.
OX NCBI_TaxID=10424;
RN [1]
RP MEDLINE=91012805; PubMed=2214027;
RX Imazeki F., Omata M., Ohto M.;
RT "Heterogeneity and evolution rates of delta virus RNA sequences.";
RL J. Virol. 64:5594-5599(1990).
CC -!- SUBCELLULAR LOCATION: Nuclear.
CC -!- PTM: Phosphorylated.
CC -!- DISEASE: BINDS SPECIFICALLY TO ANTISERA FROM PATIENTS WITH CHRONIC
CC HEPATITIS DELTA VIRAL INFECTIONS.
CC -!- MISCELLANEOUS: Delta antigen binds specifically to HDV RNA.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; D90190; BAA14214.1; -
CC F01; A36409; SAVLDM.
CC InterPro; IPR002506; HDV_ag.
CC Pfam; PF01517; HDV_ag; 1.
CC ProDom; PD002887; HDV_ag; 1.
KW Antigen; RNA-binding; Nuclear protein; Phosphorylation.
SQ SEQUENCE FROM N.A.
RX MEDLINE=91012805; PubMed=2214027;
RX Imazeki F., Omata M., Ohto M.;
RT "Heterogeneity and evolution rates of delta virus RNA sequences.";
RL J. Virol. 64:5594-5599(1990).
CC -!- SUBCELLULAR LOCATION: Nuclear.
CC -!- PTM: Phosphorylated.
CC -!- DISEASE: BINDS SPECIFICALLY TO ANTISERA FROM PATIENTS WITH CHRONIC
CC HEPATITIS DELTA VIRAL INFECTIONS.
CC -!- MISCELLANEOUS: Delta antigen binds specifically to HDV RNA.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; D90190; BAA14214.1; -
CC F01; A36409; SAVLDM.
CC InterPro; IPR002506; HDV_ag.
CC Pfam; PF01517; HDV_ag; 1.
CC ProDom; PD002887; HDV_ag; 1.
KW Antigen; RNA-binding; Nuclear protein; Phosphorylation.
SQ SEQUENCE FROM N.A.
RX MEDLINE=91012805; PubMed=2214027;
RX Imazeki F., Omata M., Ohto M.;
RT "Heterogeneity and evolution rates of delta virus RNA sequences.";
RL J. Virol. 64:5594-5599(1990).
CC -!- SUBCELLULAR LOCATION: Nuclear.
CC -!- PTM: Phosphorylated.
CC -!- DISEASE: BINDS SPECIFICALLY TO ANTISERA FROM PATIENTS WITH CHRONIC
CC HEPATITIS DELTA VIRAL INFECTIONS.
CC -!- MISCELLANEOUS: Delta antigen binds specifically to HDV RNA.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; D90191; BAA14215.1; -
CC InterPro; IPR002506; HDV_ag.
CC Pfam; PF01517; HDV_ag; 1.
CC ProDom; PD002887; HDV_ag; 1.
KW Antigen; RNA-binding; Nuclear protein; Phosphorylation.
SQ SEQUENCE 195 AA; 21799 MW; 27C708315250A442 CRC64;
Query Match 1.3%; Score 7; DB 1; Length 195;
Best Local Similarity 100.0%; Pred. No. 51;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 404 KKIKKLE 410
DB 39 KKIKKLE 45
RESULT 32
VATE PYRAB
ID VATE PYRAB STANDARD; PRT; 199 AA.
AC Q9UXU4;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE V-type ATP synthase subunit E (SC 3.6.3.14) (V-type ATPase subunit E).
GN ATP6 OR PYRAB17640 OR PAB1182.
OS Pyrococcus abyssi.
OC Archaea; Euryarchaeota; Thermococci; Thermococcales; Thermococcaceae;
OC Pyrococcus.
OX NCBI_TaxID=29292;
RN [1]
RP SEQUENCE FROM N.A.
```

```
SQ SEQUENCE 195 AA; 21811 MW; F0APP63C16F746F1 CRC64;
Query Match 1.3%; Score 7; DB 1; Length 195;
Best Local Similarity 100.0%; Pred. No. 51;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 404 KKIKKLE 410
DB 39 KKIKKLE 45
RESULT 31
AANT HDVM2
ID AANT HDVM2 STANDARD; PRT; 195 AA.
AC P25882;
DT 01-MAY-1992 (Rel. 22, Created)
DT 01-MAY-1992 (Rel. 22, Last sequence update)
DT 01-MAY-1992 (Rel. 22, Last annotation update)
DE Delta antigen.
OS Hepatitis delta virus (isolate Japanese M-2) (HDV).
OC Viruses; Deltavirus.
OX NCBI_TaxID=10425;
RN [1]
RP MEDLINE=91012805; PubMed=2214027;
RX Imazeki F., Omata M., Ohto M.;
RT "Heterogeneity and evolution rates of delta virus RNA sequences.";
RL J. Virol. 64:5594-5599(1990).
CC -!- SUBCELLULAR LOCATION: Nuclear.
CC -!- PTM: Phosphorylated.
CC -!- DISEASE: BINDS SPECIFICALLY TO ANTISERA FROM PATIENTS WITH CHRONIC
CC HEPATITIS DELTA VIRAL INFECTIONS.
CC -!- MISCELLANEOUS: Delta antigen binds specifically to HDV RNA.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; D90191; BAA14215.1; -
CC InterPro; IPR002506; HDV_ag.
CC Pfam; PF01517; HDV_ag; 1.
CC ProDom; PD002887; HDV_ag; 1.
KW Antigen; RNA-binding; Nuclear protein; Phosphorylation.
SQ SEQUENCE 195 AA; 21799 MW; 27C708315250A442 CRC64;
Query Match 1.3%; Score 7; DB 1; Length 195;
Best Local Similarity 100.0%; Pred. No. 51;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 404 KKIKKLE 410
DB 39 KKIKKLE 45
RESULT 32
VATE PYRAB
ID VATE PYRAB STANDARD; PRT; 199 AA.
AC Q9UXU4;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE V-type ATP synthase subunit E (SC 3.6.3.14) (V-type ATPase subunit E).
GN ATP6 OR PYRAB17640 OR PAB1182.
OS Pyrococcus abyssi.
OC Archaea; Euryarchaeota; Thermococci; Thermococcales; Thermococcaceae;
OC Pyrococcus.
OX NCBI_TaxID=29292;
RN [1]
RP SEQUENCE FROM N.A.
```

RC STRAIN=GES / Orsay;
RA MEDLINE=22511545; PubMed=12622808;
RA Cohen G.N., Barbe V., Plament D., Galperin M., Heilig R., Lecompte O.,
RA Poch O., Prieur D., Querellon J., Ripp R., Thierry J.-C.,
RA Van der Oost J., Weissenbach J., Zivanovic Y., Forterre P.,
RT "An integrated analysis of the genome of the hyperthermophilic
RT Archaeon Pyrococcus abyssi.";
RL Mol. Microbiol. 47:1495-1512(2003).
CC -1- FUNCTION: Produces ATP from ADP in the presence of a proton
CC gradient across the membrane.
CC -1- CATALYTIC ACTIVITY: ATP + H(2)O + H(+) (In) = ADP + phosphate +
CC H(+) (Out).
CC -1- SIMILARITY: Belongs to the V-ATPase E subunit family.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>
CC or send an email to license@isb-sib.ch).
CC
DR EMBL; AJ248288; CAB50669.1; -
DR PIR; G75028; G75028.
DR HAMAP; MF 00311; -; 1.
DR InterPro; IPR002842; ATPsynT_Esub.
DR Pfam; PF01991; V-ATP-synt_E; 1.
DR Hydrolase; ATP synthetase; Hydrogen ion transport; Complete proteome.
KW SEQUENCE 199 AA; 23132 MW; P9230F07A9D5BDE6 CRC64;
SQ
Query Match 1.3%; Score 7; DB 1; Length 199;
Best Local Similarity 100.0%; Pred. No. 52;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 346 LLKEAVE 352
DB 109 LLKEAVE 115
|||||
RESULT 33
TRPF STAPP STANDARD; PRT; 207 AA.
AC Q8CSN5;
DT 10-OCT-2003 (Rel. 42, Created)
DF 10-OCT-2003 (Rel. 42, Last sequence update)
DE N-(5'-phosphoribosyl)anthranilate isomerase (EC 5.3.1.24) (PRAI).
GN TRPF OR SE1052.
OS Staphylococcus epidermidis.
OC Bacteria; Firmicutes; Bacillales; Staphylococcus.
OX NCBI_TaxID=1282;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 12228;
RX PubMed=12950922;
RA Zhang Y.-Q., Ren S.-X., Li H.-L., Wang Y.-X., Fu G., Yang J.,
RA Qin Z.-Q., Miao Y.-G., Wang W.-Y., Chen R.-S., Shen Y., Chen Z.,
RA Yuan Z.-H., Zhao G.-P., Qu D., Danchin A., Wen Y.-M.;
RT "Genome-based analysis of virulence genes in a non-biofilm-forming
RT Staphylococcus epidermidis strain (ATCC 12228).";
RL Mol. Microbiol. 49:1577-1593(2003).
CC -1- CATALYTIC ACTIVITY: N-(5-phospho-beta-D-ribose)-anthranilate = 1-
CC (2-carboxyphenylamino)-1-deoxy-D-ribose 5-phosphate.
CC -1- PATHWAY: Tryptophan biosynthesis; third step.
CC -1- SIMILARITY: Belongs to the trpF family.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>
CC or send an email to license@isb-sib.ch).
CC

CC EMBL; AB016747; AAC04649.1; -
DR HAMAP; MF 00135; -; 1.
DR InterPro; IPR001240; PRAI.
DR Pfam; PF00697; PRAI; 1.
KW Isomerase; Tryptophan biosynthesis; Complete proteome.
SQ SEQUENCE 207 AA; 23785 MW; 7337A1DF1849582 CRC64;
Query Match 1.3%; Score 7; DB 1; Length 207;
Best Local Similarity 100.0%; Pred. No. 53;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 404 KKIKKLE 410
DB 15 KKIKKLE 21
|||||
RESULT 34
YA03 ARCFU STANDARD; PRT; 211 AA.
ID YA03 ARCFU
AC Q29259;
DT 16-OCT-2001 (Rel. 40, Created)
DF 16-OCT-2001 (Rel. 40, Last sequence update)
DE Hypothetical protein AF1003.
GN AF1003.
OS Archaeoglobus fulgidus.
OC Archaea; Euryarchaeota; Archaeoglobi; Archaeoglobales;
OC Archaeoglobaceae; Archaeoglobus.
OX NCBI_TaxID=2234;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=VC-16 / DSM 4304 / ATCC 49558;
RX MEDLINE=98049433; PubMed=9389475;
RA Klenk H.-P., Clayton R.A., Tomb J.-F., White O., Nelson K.E.,
RA Ketchum K.A., Dodson R.J., Gwinn M., Hickley E.K., Peterson J.D.,
RA Richardson D.L., Kerlavage A.R., Graham D.E., Kyrpides N.C.,
RA Fleischmann R.D., Quackenbush J., Lee N.H., Sutton G.G., Gill S.,
RA Kirkness E.F., Dougherty B.A., McKenney K., Adams M.D., Loftus B.,
RA Peterson S., Reich C.I., McNeil L.K., Badger J.H., Glöckle A., Zhou L.,
RA Overbeek R., Gocayne J.D., Weidman J.F., McDonald L., Utterback T.,
RA Cotton M.D., Spriggs T., Arliach P., Kalne B.P., Sykes S.M.,
RA Sadow P.W., D'Andrea K.P., Bowman C., Fujii C., Garland S.A.,
RA Mason T.M., Olsen G.J., Fraser C.M., Smith H.O., Woese C.R.,
RA Venter J.C.;
RT "The complete genome sequence of the hyperthermophilic, sulphate-
RT reducing archaeon Archaeoglobus fulgidus.";
RL Nature 390:364-370(1997).
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>
CC or send an email to license@isb-sib.ch).
CC
DR EMBL; AB001034; AAB90239.1; -
DR PIR; C69375; C69375.
DR TIGR; AF1003; -
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 211 AA; 24699 MW; 50411449730812F8 CRC64;
Query Match 1.3%; Score 7; DB 1; Length 211;
Best Local Similarity 100.0%; Pred. No. 54;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 165 TLSTPEE 171
DB 33 TLSTPEE 39
|||||
RESULT 35

AAANT_HDVAM
ID AAANT_HDVAM STANDARD; PRT; 214 AA.
AC P2589;
DT 01-MAY-1992 (Rel. 22, Created)
DT 01-MAY-1992 (Rel. 22, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Delta antigen.
OS Hepatitis delta virus (isolate American) (HDV).
OC Viruses; Deltavirus.
OX NCBI_TaxID=10422;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=87315446; PubMed=3627276;
RA Makino S., Chang M.F., Shieh C.K., Kamahora T., Vannier D.M.,
RA Govindarajan S., Lai M.M.C.;
RT "Molecular cloning and sequencing of a human hepatitis delta (delta)
RT virus RNA";
RL Nature 329:343-346(1987).
CC -!- SUBCELLULAR LOCATION: Nuclear.
CC -!- PTM: Phosphorylated.
CC -!- DISEASE: BINDS SPECIFICALLY TO ANTISERA FROM PATIENTS WITH CHRONIC
CC HEPATITIS DELTA VIRAL INFECTIONS.
CC -!- MISCELLANEOUS: Delta antigen binds specifically to HDV RNA.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; M28267; AA98112.1; -
CC PDB; 1A92; 08-JUN-99.
CC PDB; 1BY0; 26-SEP-01.
CC InterPro; IPR002506; HDV ag.
CC Pfam; PF01517; HDV ag; 1.
CC ProDom; PD002887; HDV ag; 1.
CC Antigen; RNA-binding; Nuclear protein; Phosphorylation; 3D-structure.
SQ SEQUENCE 214 AA; 24145 MW; 6B26DZFC0B59A50 CRC64;

Query Match 1.3%; Score 7; DB 1; Length 214;
Best Local Similarity 100.0%; Pred. No. 55;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 404 KKIKKLE 410
DB 39 KKIKKLE 45

RESULT 36
AAANT_HDVNA
ID AAANT_HDVNA STANDARD; PRT; 214 AA.
AC P2589;
DT 01-MAY-1992 (Rel. 22, Created)
DT 01-MAY-1992 (Rel. 22, Last sequence update)
DT 01-MAY-1992 (Rel. 22, Last annotation update)
DE Delta antigen.
OS Hepatitis delta virus (isolate Nauru) (HDV).
OC Viruses; Deltavirus.
OX NCBI_TaxID=10426;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=91020976; PubMed=2219700;
RA Chao Y.C., Chang M.F., Gust I., Lai M.M.C.;
RT "Sequence conservation and divergence of hepatitis delta virus RNA";
RE Virology 178:384-392(1990).
CC -!- SUBCELLULAR LOCATION: Nuclear.
CC -!- PTM: Phosphorylated.
CC -!- DISEASE: BINDS SPECIFICALLY TO ANTISERA FROM PATIENTS WITH CHRONIC
CC HEPATITIS DELTA VIRAL INFECTIONS.
CC -!- MISCELLANEOUS: Delta antigen binds specifically to HDV RNA.
CC

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; M58629; AA859753.1; -
CC PIR; A36212; SAVLDN.
CC InterPro; IPR002506; HDV ag.
CC Pfam; PF01517; HDV ag; 1.
CC ProDom; PD002887; HDV ag; 1.
CC Antigen; RNA-binding; Nuclear protein; Phosphorylation.
SQ SEQUENCE 214 AA; 24180 MW; 91F2D2B558C801D CRC64;

Query Match 1.3%; Score 7; DB 1; Length 214;
Best Local Similarity 100.0%; Pred. No. 55;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 404 KKIKKLE 410
DB 39 KKIKKLE 45

RESULT 37
SGAH_MYCFN STANDARD; PRT; 218 AA.
ID_SGAH_MYCFN STANDARD; PRT; 218 AA.
AC P75293;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Probable hexulose-6-phosphate synthase (EC 4.1.2.-) (HUMPS) (D-arabino
DE 3-hexulose 6-phosphate formaldehyde lyase).
GN SGAH OR MPN493 OR MP349.
OS Mycoplasma pneumoniae.
OC Bacteria; Firmicutes; Mollicutes; Mycoplasmataceae; Mycoplasma.
OX NCBI_TaxID=2104;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 29342 / M129;
RX MEDLINE=97105885; PubMed=8948633;
RA Himmelreich R., Hilbert H., Plagens H., Pirkel E., Li B.-C.,
RA Herrmann R.;
RT "Complete sequence analysis of the genome of the bacterium Mycoplasma
RT pneumoniae";
RL Nucleic Acids Res. 24:4420-4449(1996).
RN [2]
RP DISCUSSION OF SEQUENCE.
RX MEDLINE=97419490; PubMed=9274005;
RA Reizer J., Reizer A., Saiter M.H. Jr.;
RT "Is the ribulose monophosphate pathway widely distributed in
RT bacteria?";
RL Microbiology 143:2519-2520(1997).
CC -!- FUNCTION: Condensation of D-ribulose 5-phosphate with formaldehyde
CC to form D-arabino-6-hexulose 3-phosphate.
CC -!- PATHWAY: Probably part of a sugar metabolic pathway along with
CC sgaU and sgaV.
CC -!- SIMILARITY: BELONGS TO THE HUMPS FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; A5000033; AA895997.1; -
CC PIR; S73675; S73675.
CC InterPro; IPR001754; OMPdecase.
CC Pfam; PF00215; OMPdecase; 1.
CC Lyase; Complete proteome.
CC

SQ SEQUENCE 218 AA; 24133 MW; 5285E03858F1CB4A CRC64;
Query Match 1.3%; Score 7; DB 1; Length 218;
Best Local Similarity 100.0%; Pred. No. 56;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 264 TLNDIQL 270
Db 175 TLNDIQL 181
RESULT 38
RP29 HUMAN
ID _RP29 HUMAN STANDARD; PRT; 220 AA.
AC O95707; OSUOQ3;
DF 30-MAY-2000 (Rel. 39, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 15-MAR-2004 (Rel. 43, Last annotation update)
DE Ribonuclease P protein subunit p29 (EC 3.1.26.5) (hPOP4).
GN RPP29 OR POP4
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A., AND SEQUENCE OF 86-95.
RX MEDLINE=99146772; PubMed=10024167;
RA Jarriss N., Eder P.S., Wesolowski D., Altman S.;
RT "Rpp14 and Rpp29, two protein subunits of human ribonuclease P";
RL RNA 5:153-157(1999).
RN [2]
RP SEQUENCE FROM N.A., AND SUBCELLULAR LOCATION.
RX MEDLINE=99289489; PubMed=10352175;
RA van Bennekom H., Pruijn G.J.M., van Venrooij W.J.;
RT "hPop4: a new protein subunit of the human RNase MRP and RNase P
ribonucleoprotein complexes";
RL Nucleic Acids Res. 27:2465-2472(1999).
RN [3]
RP SEQUENCE FROM N.A.
RC TISSUE=Lung, and Lymph;
RX MEDLINE=22388257; PubMed=12477932;
RA Srausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA Klausner R.D., Collins P.S., Wagner L., Shennen C.M., Schuler G.D.,
RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
RA Dratchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
RA Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,
RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullaby S.J.,
RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Rulyk S.W.,
RA Villalón D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA Fahey J., Helton E., Kettman M., Madan A., Rodrigues S., Sanchez A.,
RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,
RA Butterfield Y.S.N., Krzywinski M.I., Skalska U., Smailus D.E.,
RA Schnarch A., Schein J.E., Jones S.J.W., Marra M.A.;
RT "Generation and initial analysis of more than 15,000 full-length
human and mouse cDNA sequences";
RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
CC -1- FUNCTION: Part of ribonuclease P, a protein complex that generates
mature tRNA molecules by cleaving their 5' ends. May function with
RPP38 to coordinate the nucleolar targeting and/or assembly of
RNase P.
CC -1- CATALYTIC ACTIVITY: Endonucleolytic cleavage of RNA, removing 5'-
extra-nucleotide from tRNA precursor.
CC -1- SUBUNIT: RNase P consists of a RNA moiety and at least 8 protein
subunits; POP1, RPP14, RPP29/POP7, RPP25, RPP29/POP4, RPP30,
RPP38 and RPP40.
CC -1- SUBCELLULAR LOCATION: Nuclear; nucleolar.
CC -1- SIMILARITY: TO YEAST POP4 AND S.POMBE SPBC1703.01C.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
or send an email to license@isb-sib.ch).
CC
CC EMBL; AF001176; AAD00893.2; -;
CC EXBL; Y18863; CAB39167.1; -;
CC EXBL; BC004438; AAO04438.1; -;
CC EXBL; BC006098; AAH06098.1; -;
CC GK; O95707; -;
CC MTM; 606114; -;
CC GO; GO:0000172; C:ribonuclease MRP complex; TAS.
CC GO; GO:0004526; F:ribonuclease P activity; TAS.
CC GO; GO:0003723; F:RNA binding; TAS.
CC GO; GO:0006364; P:RNA processing; TAS.
CC GO; GO:0008033; P:RNA processing; TAS.
CC InterPro; IPR002730; UPF0086.
CC Pfam; PF01868; UPF0086; 1.
CC SMART; SM00538; POP4; 1.
CC Hydrolase; Nuclear protein; tRNA processing.
KW DOMAIN 65 74 POLY-DYS.
FT CONFLICT 38 39 ST -> TS (IN REF. 1).
SQ SEQUENCE 220 AA; 25424 MW; 26141947B9F5C61E CRC64;
Query Match 1.3%; Score 7; DB 1; Length 220;
Best Local Similarity 100.0%; Pred. No. 56;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 146 KKKAKGL 152
Db 70 KKKAKGL 76
RESULT 39
YNCC _SCOLI
ID YNCC _SCOLI STANDARD; PRT; 221 AA.
AC P76114; P77347;
DT 15-JUL-1999 (Rel. 38, Created)
DT 15-JUL-1999 (Rel. 38, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Hypothetical transcriptional regulator yncC.
GN YNCC OR B1450.
OS Escherichia coli.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Escherichia.
OX NCBI_TaxID=562;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=K12 / MG1655;
RX MEDLINE=97426617; PubMed=9278503;
RA Blattner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,
RA Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,
RA Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,
RA Mau B., Shao Y.;
RT "The complete genome sequence of Escherichia coli K-12.";
RL Science 277:1453-1474(1997).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=K12;
RX MEDLINE=97251357; PubMed=9097039;
RA Aiba H., Baba T., Fujita K., Hayashi K., Inada T., Isono K., Itoh T.,
RA Kasai H., Kashimoto K., Kimura S., Kitakawa M., Kitagawa M.,
RA Makino K., Miki T., Mizobuchi K., Mori H., Mori T., Motomura K.,
RA Nakade S., Nakamura Y., Nishimoto H., Nishio Y., Oshima T., Saito N.,
RA Sampaio G., Seki Y., Sivasubramanian S., Tagami H., Takeda J.,
RA Takemoto K., Takeuchi Y., Wada C., Yamamoto Y., Horiuchi T.;
RT "A 570-kb DNA sequence of the Escherichia coli K-12 genome
corresponding to the 28.0-40.1 min region on the linkage map";
RL DNA Res. 3:363-377(1996).
CC -1- SIMILARITY: Belongs to the gntR family of transcriptional

```

CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch)
CC -----
CC EMBL: AE000242; AAC74532.1; ALT_INIT.
CC EMBL: D90784; BAA15082.1; -.
CC EMBL: D90785; BAA15085.1; -.
CC EcoGene; EG13773; yncC.
CC InterPro; IPR000524; HTH_GntR.
CC Pfam; PF003192; GntR; 1.
CC SMART; SM00345; HTH_GNTR; 1.
CC PROSITE; PS00043; HTH_GNTR_FAMILY; FALSE NEG.
KW Hypothetical protein; Transcription regulation; DNA-binding;
FT DNA BIND 37 56 H-T-H MOTIF (POTENTIAL).
SQ SEQUENCE 221 AA; 25151 MW; 7A589BCFAE20D75C CRC64;

Query Match 1.3%; Score 7; DB 1; Length 221;
Best Local Similarity 100.0%; Pred. No. 57;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 324 KIQQAQ 330
Db 118 KIQQAQ 124

RESULT 40
TLIY ARATH
ID TLIY ARATH STANDARD; PRT; 225 AA.
AC Q9SW33;
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Thylakoid lumenal 17.9 kDa protein, chloroplast precursor.
GN AtG24930 OR F13M23.70.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; Core eudicots; rosids;
OC euroids II; Brassicales; Brassicaceae; Arabidopsi.
OX NCBI_TaxID=3702;
[1]
SEQUENCE FROM N.A.
RP STRAIN=sv. Columbia;
RX MEDLINE=20083498; PubMed=10617198;
RA Mayer K.F.X., Schueller A., Stiekema W., Entian K.-D., Terryn N.,
RA Pohl T., Duesterhoeft A., Stiekema W., Entian K.-D., Terryn N.,
RA Harris B., Anstorge W., Brandt P., Grivell L.A., Rieger M.,
RA Weichselgartner M., de Simone V., Obermaier B., Mache R., Mueller M.,
RA Kreis M., Delsen M., Puigdomenech P., Watson M., Schmidtheini T.,
RA Reichert B., Portetelle D., Perez-Alonso M., Boutry M., Bancroft I.,
RA Vos P., Heiseil J., Zimmermann W., Wedler H., Ridley P.,
RA Langham S.-A., McCullagh B., Bilham L., Robben J.,
RA Van der Schueren J., Grynoprez B., Chuang Y.-J., Vandenbussche F.,
RA Braeken M., Weltjens I., Voet M., Bastiaens I., Aert R., Defoor E.,
RA Weitzenegger T., Bothe G., Ransperger U., Hilbert H., Braun M.,
RA Holzer E., Brandt A., Peters S., van Staveren M., Dirkse W.,
RA Moolijman P., Klein Lankhorst R., Rose M., Hauf J., Koetter P.,
RA Berreiser S., Hempel S., Feldpausch M., Lamberth S., Van den Daele H.,
RA De Keyser A., Buysshaert C., Gielen J., Villarroel R., De Clercq R.,
RA Van Montagu M., Rogers J., Cronin A., Quail M.A., Bray-Allen S.,
RA Clark L., Doggett J., Hall S., Kay M., Lennard N., McLay K., Mayes R.,
RA Pettett A., Rajandream M.A., Lyne M., Benes V., Reckmann S.,
RA Borkova D., Bloeker H., Scharfe M., Grimm M., Loehner T.-H.,
RA Dose S., de Haan M., Maarse A.C., Schaefer M., Mueller-Auer S.,
RA Gabel C., Fuchs M., Fartmann B., Granderath K., Dauner D., Herzl A.,
RA Neumann S., Argiriou A., Vitale D., Ligouri R., Piravandi E.,
RA Massenot O., Quigley F., Clabaud G., Muendlein A., Felber R.,

```

```

RA Schnabl S., Hiller R., Schmidt W., Lecharny A., Aubourg S.,
RA Chedor P., Cooke R., Berger C., Monfort A., Casacuberta E.,
RA Gibbons T., Weber N., Vandenbol M., Baigues M., Terol J., Torres A.,
RA Perez-Perez A., Purnelle B., Bent E., Johnson S., Tacon D., Jesse T.,
RA Heijnen L., Schwarz S., Scholler P., Heber S., Francs P., Biele C.,
RA Frishman D., Haase D., Lemcke K., Mewes H.-W., Stocker S.,
RA Zaccaria P., Bevan M., Wilson R.K., de la Bastide M., Habermann K.,
RA Parnell L., Dedhia N., Gnoj L., Schutz K., Huang E., Spiegel L.,
RA Sakhon M., Murray J., Sheet P., Cordes M., Abu-Threideh J.,
RA Stenking T., Kalicki J., Graves T., Harmon G., Edwards J.,
RA Lestraille P., Courtney L., Cloud J., Abbott A., Scott K., Johnson D.,
RA Minx P., Bentley D., Fulton B., Miller N., Greco T., Kemp K.,
RA Kramer J., Fulton L., Mardis E., Dante M., Pepin K., Hillier L.W.,
RA Nelson J., Spieth J., Ryan E., Andrews S., Geisel C., Layman D.,
RA Du H., Ali J., Berghoff A., Jones K., Drone K., Cotton M., Joshi C.,
RA Antoniou B., Zidanic M., Strong C., Sun H., Lamar B., Yordan C.,
RA Ma P., Zhong J., Preston R., Vill D., Shekher M., Matero A., Shah R.,
RA Swaby I.K., O'Shaughnessy A., Rodriguez M., Hoffman J., Till S.,
RA Grant S., Shohdy N., Hasegawa A., Hameed A., Lodhi M., Johnson A.,
RA Chen E., Marra M.A., Martienssen R., McCombie W.R.;
RT "Sequence and analysis of chromosome 4 of the plant Arabidopsis
RT thaliana."
RL Nature 402:769-777(1999).
RN [2]
RP SEQUENCE OF 64-93;
RC STRAIN=sv. Columbia;
RA Schubert M., Peterson U., Funk C., Schroeder W.P., Kieselbach T.;
RL Submitted (AUG-2001) to Swiss-Prot.
CC -1- SUBCELLULAR LOCATION: Chloroplast; within the thylakoid lumen.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch)
CC -----
CC EMBL: AL035523; CAB36735.1; -.
CC EMBL: AL161562; CAB79402.1; -.
CC FRL; T05514; T05514.
CC TRANSIT 1 ? CHLOROPLAST (POTENTIAL).
CC TRANSIT 1 ? THYLAKOID.
CC TRANSIT 1 ? THYLAKOID.
CC CHAIN 64 225 THYLAKOID LUMENAL 17.9 KDA PROTEIN.
CC SEQUENCE 225 AA; 24695 MW; 067C3FEDC1FP4C41 CRC64;

Query Match 1.3%; Score 7; DB 1; Length 225;
Best Local Similarity 100.0%; Pred. No. 58;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 291 ERLKKLI 297
Db 205 ERLKKLI 211

RESULT 41
PYRH AERPE
ID PYRH AERPE STANDARD; PRT; 227 AA.
AC Q9YF40;
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Probable uridylylate kinase (EC 2.7.4.-) (UK) (Uridine monophosphate
DE kinase) (UMP kinase).
GN PYRH OR APB0401.
OS Aeropyrum pernix.
OC Archaea; Crenarchaeota; Thermoprotei; Desulfurococcales;
OC Desulfurococaceae; Aeropyrum.
OX NCBI_TaxID=56636;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=X1;

```

```

RX MEDLINE=99310339; PubMed=10382966;
RA Kawarabayashi Y., Hino Y., Horikawa H., Yamazaki S., Haikawa Y.,
RA Jin-no K., Takahashi M., Sekine M., Baba S.-I., Ankai A., Kosugi H.,
RA Hosoyama A., Fukui S., Nagai Y., Nishijima K., Nakazawa H.,
RA Takamiya M., Masuda S., Funahashi T., Tanaka T., Kudoh Y.,
RA Yamazaki J., Kishida N., Ouchi A., Aoki K.-I., Kubota K.,
RA Nakamura Y., Nomura N., Sako Y., Kikuchi H.,
RT "complete genome sequence of an aerobic hyperthermophilic
RL crenarchaeon, Aeropyrum pernix K1.";
CC DNA Res. 6:93-101(1999).
CC -|- FUNCTION: Catalyzes the phosphorylation of UMP to UDP (by
CC similarity).
CC -|- CATALYTIC ACTIVITY: ATP + UMP = ADP + UDP.
CC -|- PATHWAY: Pyrimidine biosynthesis; conversion of UMP to CTP; first
CC step.
CC -|- SIMILARITY: Belongs to the UMP kinase family.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC -----
DR EMBL; AP000059; BAA79356.1; ALT_INIT.
DR HAMAP; MF_01220; -; 1.
DR InterPro; IPR001048; Aa_kinase.
DR Pfam; PF00636; aakinas; 1.
KW Transferrase; Kinase; Pyrimidine biosynthesis; Complete proteome.
SQ SEQUENCE 227 AA; 24400 MW; C1691C26A240FA6A CRC64;

Query Match 1.3%; Score 7; DB 1; Length 227;
Best Local Similarity 100.0%; Pred. No. 58;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 61 RQEDIL 67
Db 167 RQEDIL 173

RESULT 42
REGO_BP82 STANDARD; PRT; 229 AA.
AC P13870;
DT 01-JAN-1990 (Rel. 13, Created)
DT 01-JAN-1990 (Rel. 13, Last sequence update)
DT 15-MAR-2004 (Rel. 43, Last annotation update)
DE Antitermination protein Q.
GN Q.
OS Bacteriophage 82.
OC Viruses; dsDNA viruses, no RNA stage; Caudovirales; Siphoviridae;
OC Lambda-like viruses.
OC NCBI_TaxID=10705;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=87308148; PubMed=3624233;
RA Golliger J.A., Roberts J.W.;
RT "Bacteriophage 82 gene Q and Q protein. Sequence, overproduction, and
RT activity as a transcription antiterminator in vitro.";
RL J. Biol. Chem. 262:11721-11725(1987).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=96196428; PubMed=8648624;
RA Mandi A.A., Sharples G.J., Mandal T.N., Lloyd R.G.;
RT "Holliday junction resolvases encoded by homologous rusa genes in
RT Escherichia coli K-12 and phage 82.";
RL J. Mol. Biol. 257:561-573(1996).
CC -|- FUNCTION: Positively regulates expression of the phage late gene
CC operons. Bacterial host RNA polymerase modified by antitermination
CC proteins transcribes through termination sites that otherwise
CC prevent expression of the regulated genes.
CC
CC -----

```

```

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC -----
DR EMBL; J02803; AAA32298.1; -.
DR EMBL; X92588; CAA63332.1; -.
DR PIR; A29791; PQBP82.
KW Transcription regulation; Transcription termination; DNA-binding.
FT DNA BIND 14 32 POTENTIAL.
SQ SEQUENCE 229 AA; 26400 MW; E6EDABE818498667 CRC64;

Query Match 1.3%; Score 7; DB 1; Length 229;
Best Local Similarity 100.0%; Pred. No. 58;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 291 ERLKKLI 297
Db 144 ERLKKLI 150

RESULT 43
MYPO_HUMAN STANDARD; PRT; 248 AA.
AC P25189; Q16072; Q92677; Q9BR67;
DT 01-MAY-1992 (Rel. 22, Created)
DT 01-MAY-1992 (Rel. 22, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Myelin P0 protein precursor (Myelin protein zero) (Myelin peripheral
DE protein) (MPP).
DE GN MPZ.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OC NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=92062068; PubMed=1719967;
RA Hayasaka K., Nanao K., Tahara M., Sato W., Takada G., Miura M.,
RA Uemura K.;
RT "Isolation and sequence determination of cDNA encoding the major
RT structural protein of human peripheral myelin.";
RL Biochem. Biophys. Res. Commun. 180:515-518(1991).
RN [2]
RP SEQUENCE FROM N.A., AND VARIANT CMT1B HTS-98.
RC TISSUE=Spinal cord;
RX MEDLINE=93356807; PubMed=7688964;
RA Hayasaka K., Ohnishi A., Takada G., Fukushima Y., Murai Y.;
RT "Mutation of the myelin P0 gene in Charcot-Marie-Tooth neuropathy
RT type 1.";
RL Biochem. Biophys. Res. Commun. 194:1317-1322(1993).
RN [3]
RP SEQUENCE FROM N.A.
RX MEDLINE=94154677; PubMed=7509228;
RA Pham-Dinh D., Fourbil Y., Blanguet F., Mattei M.-G., Roeckel N.,
RA Latour P., Chazot G., Vandenbergh A., Dautigny A.;
RT "The major peripheral myelin protein zero gene: structure and
RT localization in the cluster of Fc gamma receptor genes on human
RT chromosome 1q21.3-q23.";
RL Hum. Mol. Genet. 2:2051-2054(1993).
RN [4]
RP SEQUENCE FROM N.A.
RX MEDLINE=96196428; PubMed=8648624;
RA Mandi A.A., Sharples G.J., Mandal T.N., Lloyd R.G.;
RT "Holliday junction resolvases encoded by homologous rusa genes in
RT Escherichia coli K-12 and phage 82.";
RL J. Mol. Biol. 257:561-573(1996).
CC -|- FUNCTION: Positively regulates expression of the phage late gene
CC operons. Bacterial host RNA polymerase modified by antitermination
CC proteins transcribes through termination sites that otherwise
CC prevent expression of the regulated genes.
CC
CC -----

```

RC TISSUE=Skin;
RX MEDLINE=22398257; PubMed=12477932;
RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hong L.,
Datchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
Rana S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullany S.J.,
Bosak S.A., McWain P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
Richards S., Worley K.C., Hale S.J., Garcia A.M., Gay L.J., Hulyk S.W.,
Villalon D.K., Muzny N.J., Sedgwick E.J., Lu X., Gibbs R.A.,
Fahey J., Helton E., Kettner M., Madan A., Rodriguez S., Sanchez A.,
Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,
Butterfield Y.S.N., Krzywinski M.I., Skalska U., Smalley D.B.,
Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.,
RT "Generation and initial analysis of more than 15,000 full-length human
and mouse cDNA sequences.";
RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
RN [6]
RP SEQUENCE OF 24-248 FROM N.A.
RA Nelis E., Timmerman V., De Jonghe P., Mylly L., Martin J.-J.,
Van Broeckhoven C.;
RT "Linkage and mutation analysis in an extended family with Charcot-
Marie-Tooth disease type 1B.";
RL Submitted (MAY-1994) to the EMBL/GenBank/DBJ databases.
RN [7]
RP REVIEW ON CMT1B VARIANTS.
RX MEDLINE=95282670; PubMed=7762451;
RA Roa B.B., Lupski J.R.;
RT "Molecular Genetics of Charcot-Marie-Tooth neuropathy.";
RL Adv. Hum. Genet. 22:117-152(1994).
RN [8]
RP REVIEW ON CMT1B VARIANTS.
RX MEDLINE=94302675; PubMed=7518101;
RA Patel P.I., Lupski J.R.;
RT "Charcot-Marie-Tooth disease: a new paradigm for the mechanism of
inherited disease.";
RL Trends Genet. 10:128-133(1994).
RN [9]
RP REVIEW ON CMT1B AND DSS VARIANTS.
RX MEDLINE=99103460; PubMed=9889385;
RA Nelis E., Haates N., van Broeckhoven C.;
RT "Mutations in the peripheral myelin genes and associated genes in
inherited peripheral neuropathies.";
RL Hum. Mutat. 13:11-28(1999).
RN [10]
RP VARIANT CMT1B MET-30.
RX MEDLINE=94061030; PubMed=7694726;
RA Hayasaka K., Takada G., Ionescu V.V.;
RT "Mutation of the myelin P0 gene in Charcot-Marie-Tooth neuropathy
type 1B.";
RL Hum. Mol. Genet. 2:1369-1372(1993).
RN [11]
RP VARIANT CMT1B CVS-82.
RX MEDLINE=94083941; PubMed=7505151;
RA Himoro M., Yoshikawa H., Matsui T., Mitsui Y., Takahashi M.,
Kado M., Nishimura T., Sawashita Y., Takada G., Hayasaka K.;
RT "New mutation of the myelin P0 gene in a pedigree of
Charcot-Marie-Tooth neuropathy 1.";
RL Biochem. Mol. Biol. Int. 31:169-173(1993).
RN [12]
RP VARIANTS CMT1B GLU-90 AND GLU-96.
RX MEDLINE=94035113; PubMed=7693129;
RA Hayasaka K., Himoro M., Sato W., Takada G., Uyemura K., Shimizu N.,
Bird T.D., Conneally P.M., Chance P.F.;
RT "Charcot-Marie-Tooth neuropathy type 1B is associated with mutations
of the myelin P0 gene.";
RL Nat. Genet. 5:31-34(1993).
RN [13]

RP VARIANTS CMT1B LEU-78 AND ASN-101.
RX MEDLINE=96055517; PubMed=7550231;
RA Latour P., Blanquet F., Nelis E., Bonnebouche C., Chapon F.,
Diraison P., Ollagnon E., Dautigny A., Pham-Dinh D., Chazot G.,
Boucherat M., van Broeckhoven C., Vandenbergh A.;
RT "Mutations in the myelin protein zero gene associated with
Charcot-Marie-Tooth disease type 1B.";
RL Hum. Mutat. 6:50-54(1995).
RN [19]
RP VARIANTS CMT1B PHE-64 DEL.
RX MEDLINE=96212920; PubMed=8630052;
RA Ikegami T., Nicholson G.A., Ikeda H., Ishida A., Johnston H., Wise G.,
Ouvrier R.A., Hayasaka K.;
RT "A novel homozygous mutation of the myelin P0 gene producing
Dejerine-Sottas disease (hereditary motor and sensory neuropathy
type III).";
RL Biochem. Biophys. Res. Commun. 222:107-110(1996).
RN [20]
RP VARIANTS CMT1B THR-135 AND SER-137.
RX MEDLINE=96253736; PubMed=8664899;
RA Roa B.B., Warner L.E., Garcia C.A., Russo D., Lovelace R.,
Chance P.F., Lupski J.R.;
RT "Myelin protein zero (MPZ) gene mutations in nonduplication type 1
Charcot-Marie-Tooth disease.";
RL Hum. Mutat. 7:36-45(1996).
RN [21]
RP VARIANTS CMT1B SER-122.
RX MEDLINE=97001227; PubMed=8844219;
RA Blanquet-Grossard F., Pham-Dinh D., Dautigny A., Latour P.,
Bonnebouche C., Diraison P., Chapon F., Chazot G., Vandenbergh A.;
RT "Charcot-Marie-Tooth type 1B neuropathy: a mutation at the single
glycosylation site in the major peripheral myelin glycoprotein P0.";
RN [22]

AC Q57083;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 15-MAR-2004 (Rel. 43, Last annotation update)
DE HTH-type transcriptional regulator perR (Peroxide resistance protein
perR).
DE perR.
GN PER OR B0254.
OS Escherichia coli.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Escherichia.
OX NCBI_TaxID=562;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=K12;
RA Loewen P.C., Swisala J., Volkert M.R.;
RL Submitted (JUN-1996) to the EMBL/GenBank/DBSJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=K12 / MG1655;
RX MEDLINE=97426617; PubMed=9278503;
RA Blattner P.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,
RA Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.P.,
RA Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,
RA Mau B., Shao Y.;
RT "The complete genome sequence of Escherichia coli K-12.";
RL Science 277:1453-1474[1997].
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=K12 / W3110;
RA Takemoto K., Mori H., Murayama N., Kataoka K., Yano M., Itoh T.,
RA Yamamoto Y., Inokuchi H., Miki T., Hatada E., Fukuda R., Ichihara S.,
RA Mizuno T., Makino K., Nakata A., Yura T., Sampaï G., Mizobuchi K.;
RT "Systematic sequencing of the Escherichia coli genome: analysis of the
RT 4.0 - 6.0 min (189,987 - 281,416bp) region.";
RL Submitted (FEB-1996) to the EMBL/GenBank/DBSJ databases.
RN [4]
RP SEQUENCE FROM N.A.
RA Schramm S., Duncan M., Allen E., Araujo R., Aparicio A., Chung E.,
RA Davis K., Pederspiel N., Hyman R., Kalman S., Komp C., Kurdi O.,
RA Lashkari D., Lew H., Lin D., Namath A., Oefner P., Roberts D.,
RA Davis R.W.;
RL Submitted (SEP-1996) to the EMBL/GenBank/DBSJ databases.
CC -!- FUNCTION: APPARENT REGULATORY GENE INVOLVED IN PEROXIDE RESISTANCE
CC IN STATIONARY PHASE.
CC -!- SIMILARITY: Contains 1 HTH lyseR-type DNA-binding domain.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; U57080; AAB01994.1; -.
CC EMBL; AE000133; RAC73357.1; -.
CC EMBL; D83336; BAA77324.1; -.
CC EMBL; U70214; AAB08673.1; -.
CC PIR; F64750; F64750.
CC EcoGene; EG13340; perR.
CC InterPro; IPR000847; HTH LysR.
CC InterPro; IPR005119; LysR_subst.
CC Pfam; PF00126; HTH_1; 1.
CC Pfam; PF03466; LysR_substrate; 1.
CC PRINTS; PRO0039; HTHLYSR.
CC PROSITE; PS50931; HTH_LYSR; 1.
CC Transcription regulation; DNA-binding; Complete proteome.
FT DOMAIN 7 64 HTH LYSR-TYPE.
FT DNA_BIND 24 44 H-T-H MOTIF (POTENTIAL).
SQ SEQUENCE 297 AA; 33638 MW; E197CE2334B7420E CRC64;
Query Match 1.3%; Score 7; DB 1; Length 297;
Best Local Similarity 100.0%; Pred. No. 73;

Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 289 LAERLKK 295
Db 173 LAERLKK 179
RESULT 46
MURB CLOPE STANDARD; PRT; 304 AA.
AC Q8XNIO;
DT 10-OCT-2003 (Rel. 42, Created)
DT 10-OCT-2003 (Rel. 42, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE UDP-N-acetylglucosaminylglucosamine reductase (EC 1.1.1.158) (UDP-N-
DE acetylglucosamine dehydrogenase).
GN MURB OR CPE0353.
OS Clostridium perfringens.
OC Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;
OC Clostridium.
OX NCBI_TaxID=1502;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=13 / Type A;
RX MEDLINE=21664373; PubMed=1192842;
RA Shimizu T., Ohtani K., Hirakawa H., Ohshima K., Yamashita A.,
RA Shiba T., Ogasawara N., Hattori M., Kuhara S., Hayashi H.;
RT "Complete genome sequence of Clostridium perfringens, an anaerobic
RT flesh-eater.";
RL Proc. Natl. Acad. Sci. U.S.A. 99:996-1001(2002).
CC -!- FUNCTION: Cell wall formation (by similarity).
CC -!- CATALYTIC ACTIVITY: UDP-N-acetylmuramate + NADP(+) = UDP-N-
CC acetyl-3-O-(1-carboxyvinyl)-D-glucosamine + NADPH.
CC -!- COFACTOR: PAD (by similarity).
CC -!- PATHWAY: Peptidoglycan biosynthesis.
CC -!- SUBCELLULAR LOCATION: Cytoplasmic (Probable).
CC -!- SIMILARITY: Belongs to the murB family.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; AP003186; BAB80059.1; -.
CC HAMAP; MF 00037; -; 1.
CC InterPro; IPR003170; MurB.
CC InterPro; IPR006094; Oxid_PAD bind.
CC Pfam; PF01565; PAD_binding_4; 1.
CC Pfam; PF02873; MurB_C; 1.
CC Oxidoreductase; NADP; Flavoprotein; PAD; Cell wall; Cell division;
CC Peptidoglycan synthesis; Complete proteome.
SQ SEQUENCE 304 AA; 33267 MW; CB4C78C107A6594B CRC64;
Query Match 1.3%; Score 7; DB 1; Length 304;
Best Local Similarity 100.0%; Pred. No. 75;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 46 TAQSGAL 52
Db 101 TAQSGAL 107
RESULT 47
Y280 ARATH STANDARD; PRT; 308 AA.
ID Y280 ARATH
AC Q82368;
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Hypothetical protein At2g29880.

GN AT2G39880 OR F6K5.1
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids;
OC eurosids II; Brassicales; Brassicaceae; Arabidopsids.
OX NCBI_TaxID=3702;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=cv. Columbia;
RX MEDLINE=20083487; PubMed=106171797;
RA Lin X., Kaul S., Rounsley S.D., Shea T.P., Benito M.-I., Town C.D.,
RA Fujii C.Y., Mason T.M., Bowman C.L., Barnstead M.E., Feldblum T.V.,
RA Buell C.R., Ketchum K.A., Lee J.J., Ronning C.M., Koo H.L.,
RA Moffat K.S., Cronin L.A., Shen M., Pai G., Van Aken S., Unayam L.,
RA Tallon L.J., Gil J.E., Adams M.D., Carrera A.J., Creasy T.H.,
RA Goodman H.M., Somerville C.R., Copenhaver G.P., Preuss D.,
RA Niernan W.C., White O., Eisen J.A., Salzberg S.L., Fraser C.M.,
RA Venter J.C.;
RT "Sequence and analysis of chromosome 2 of the plant Arabidopsis
RT thaliana";
RL Nature 402:761-768(1999).
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AC007113; AB23628.1;
DR EMBL; AC005496; AA015055.1;
DR PIR; G84701; G84701.
KW Hypothetical protein.
SQ SEQUENCE 308 AA; 35622 MW; 9C8D18E273C09BES CRC64;

Query Match 1.3%; Score 7; DB 1; Length 308;
Best Local Similarity 100.0%; Pred. No. 76;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 336 EERHORE 342
Db 242 EERHORE 248

RESULT 48
ID HTPX THEAC STANDARD; PRT; 317 AA.
AC Q9HJV2;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Probable protease htpx homolog (EC 3.4.24.-).
GN HTPX OR TA0861.
OS Thermoplasma acidophilum.
OC Archaea; Euryarchaeota; Thermoplasmata; Thermoplasmatales;
OC Thermoplasmataceae; Thermoplasma.
OX NCBI_TaxID=2303;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=DSM 1728;
RX MEDLINE=20479972; PubMed=11029001;
RA Ruepp A., Graml W., Santos-Martinez M.-L., Koretke K.K., Volker C.,
RA Mewes H.-W., Frishman D., Stocker S., Lupas A.N., Baumeister W.;
RT "The genome sequence of the thermoacidophilic scavenger Thermoplasma
RT acidophilum";
RL Nature 407:508-513(2000).
CC -!- COFACTOR: Binds 1 zinc ion per subunit (By similarity).
CC -!- SUBCELLULAR LOCATION: Integral membrane protein (potential).
CC -!- SIMILARITY: Belongs to peptidase family M48.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -

CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AL445065; CAC11990.1;
DR HAMAP; MF 00188; -; 1.
DR InterPro; IPR006025; Pept_M_Zn_BS.
DR InterPro; IPR001915; Peptidase M48.
DR Pfam; PF01435; Peptidase M48; 1.
DR PROSITE; PS00142; ZINC_PROTEASE; FALSE_NEG.
KW Transmembrane; Hydrolase; Metalloprotease; Zinc; Complete proteome.
FT TRANSMEM 14 34 POTENTIAL.
FT TRANSMEM 41 61 POTENTIAL.
FT TRANSMEM 158 178 POTENTIAL.
FT TRANSMEM 189 209 POTENTIAL.
FT METAL 146 146 ZINC (CATALYTIC) (BY SIMILARITY).
FT ACT_SITE 147 147 ZINC (CATALYTIC) (BY SIMILARITY).
FT METAL 150 150 ZINC (CATALYTIC) (BY SIMILARITY).
SQ SEQUENCE 317 AA; 35616 MW; 535EFBA7F73CDEE CRC64;

Query Match 1.3%; Score 7; DB 1; Length 317;
Best Local Similarity 100.0%; Pred. No. 77;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 448 IQRLEKL 454
Db 308 IQRLEKL 314

RESULT 49
ID THIL METJA STANDARD; PRT; 319 AA.
AC Q60337;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Probable thiamine-monophosphate kinase (EC 2.7.4.16) (Thiamine-
DE phosphate kinase).
GN THIL OR MJ0028.
OS Methanococcus jannaschii.
OC Archaea; Euryarchaeota; Methanococci; Methanococcales;
OC Methanocaldococaceae; Methanocaldococcus.
OX NCBI_TaxID=2190;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=JAL-1 / DSM 2661 / ATCC 43067;
RX MEDLINE=96337999; PubMed=8688087;
RA Bult C.J., White O., Olsen G.J., Zhou L., Fleischmann R.D.,
RA Sutton G.G., Blake J.A., FitzGerald L.M., Clayton R.A., Gocayne J.D.,
RA Kerlavage A.R., Dougherty B.A., Tomb J.-P., Adams M.D., Reich C.I.,
RA Overbeek R., Kirkness E.F., Weinstock K.G., Merrick J.M., Glodek A.,
RA Scott J.L., Geoghegan N.S.M., Weidman J.F., Fuhrmann J.L., Nguyen D.,
RA Uitterback T.R., Kelley J.M., Peterson J.D., Sadow P.W., Hanna M.C.,
RA Cotton M.D., Roberts K.M., Hurst M.A., Kaine B.P., Borodovsky M.,
RA Klenk H.-P., Fraser C.M., Smith H.O., Woese C.R., Venter J.C.;
RT "Complete genome sequence of the methanogenic archaeon, Methanococcus
RT jannaschii";
RL Science 273:1058-1073 (1996).
CC -!- CATALYTIC ACTIVITY: ATP + thiamine phosphate = ADP + thiamine
CC diphosphate.
CC -!- PATHWAY: Thiamine biosynthesis.
CC -!- SIMILARITY: Belongs to the thiamine-monophosphate kinase family.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----


```
CC EMBL; U67461; AAB98008.1; -.
DR F1R; D64303; D64303.
DR TIGR; MJO028; -.
DR InterPro; IPR000728; AIR synth.
DR InterPro; IPR006283; ThiL.
DR Pfam; PF00386; AIRS; 1.
DR TIGRFAMs; TIGR01379; ThiL; 1.
DR Thiamine biosynthesis; transferase; Kinase; Complete proteome.
KW KW
SQ SEQUENCE 319 AA; 36160 MW; 92607DSAD59A723F CRC64;

Query Match 1.3%; Score 7; DB 1; Length 319;
Best Local Similarity 100.0%; Pred. No. 78;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 151 GLGKEIT 157
DB 222 GLGKEIT 228

RESULT 50
RPOA_UREPA
ID RPOA_UREPA STANDARD; PRT; 327 AA.
AC Q9P0N4;
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE DNA-directed RNA polymerase alpha chain (EC 2.7.7.6) (RNAP alpha subunit) (Transcriptase alpha chain) (RNA polymerase alpha subunit).
GN RPOA OR UO257.
OS Ureaplasma parvum (Ureaplasma urealyticum biotype 1).
OC Bacteria; Firmicutes; Mollicutes; Mycoplasmataceae; Ureaplasma.
OX NCBI_TaxID=134821;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Setovar 3;
RX MEDLINE=20500219; PubMed=11048724;
RA Glass J.I., Lefkowitz E.J., Glass J.S., Heiner C.R., Chen E.Y., Caswell G.H.;
RT "The complete sequence of the mucosal pathogen Ureaplasma urealyticum."
RL Nature 407:757-762(2000).
CC -!- FUNCTION: DNA-dependent RNA polymerase catalyzes the transcription of DNA into RNA using the four ribonucleoside triphosphates as substrates.
CC -!- SUBUNIT: Nucleoside triphosphate = N diphosphate + [RNA] (N).
CC -!- SUBUNIT: Homodimer. The RNAP catalytic core consists of 2 alpha, 1 beta, 1 beta', and 1 omega subunit (By similarity).
CC -!- DOMAIN: The N-terminal domain is essential for RNAP assembly and basal transcription, whereas the C-terminal domain is involved in interaction with transcriptional regulators and with upstream promoter elements (By similarity).
CC -!- SIMILARITY: Belongs to the RNA polymerase alpha chain family.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See http://www.ebi.ac.uk/announcements/ or send an email to license@ebi.ac.uk).
CC -----
CC EMBL; AE002122; AAF30666.1; -.
CC HSP; P00574; ICOO.
CC HAVAP; MF_000599; -.
CC InterPro; IPR009025; RBP11-like RNAPo.
CC InterPro; IPR001700; RNA polA_bac_org.
CC Pfam; PF01000; RNA pol A bac; 1.
CC Pfam; PF03118; RNA pol A bac; 1.
CC ProDom; PD001179; RNA polA_bac_org; 2.
CC SMART; SM00642; RPOD; 1.
KW Transferase; Transcription; DNA-directed RNA polymerase;
```

```
KW Complete proteome.
FT DOMAIN 1 242 ALPHA N-TERMINAL DOMAIN (ALPHA-NTD)
FT 259 327 (BY SIMILARITY).
FT DOMAIN 259 327 ALPHA C-TERMINAL DOMAIN (ALPHA-CTD)
FT 327 327 (BY SIMILARITY).
SQ SEQUENCE 327 AA; 36817 MW; 4BAEC583237902E4 CRC64;

Query Match 1.3%; Score 7; DB 1; Length 327;
Best Local Similarity 100.0%; Pred. No. 80;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 248 REEEKR 254
DB 254 REEEKR 260

RESULT 51
IF2B_MOUSE
ID IF2B_MOUSE STANDARD; PRT; 331 AA.
AC Q9SLA5; O9CSH6; O9CT12;
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 15-MAR-2004 (Rel. 43, Last annotation update)
DE Eukaryotic translation initiation factor 2 subunit 2 (Eukaryotic translation initiation factor 2 beta subunit) (eIF-2-beta).
GN EIF2S2.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=Embryo;
RX MEDLINE=22354683; PubMed=12466851;
RA Nikaido I., Furuno M., Kasukawa T., Adachi J., Bono H., Kondo S., Okazaki Y., Tomaru Y., Hasegawa Y., Nogami A., Schonbach C., Gojbori T., Baldarelli R., Hill D.P., Bult C., Hume D.A., Quackenbush J., Schiraldi L.M., Kanapin A., Matsuda H., Batalov S., Beisel K.W., Blake J.A., Bradt D., Brusic V., Chothia C., Corbani L.E., Cousins S., Dalla B., Dragani T.A., Fletcher C.P., Forrest A., Frazer K.S., Gaasterland T., Gariboldi M., Gissi C., Godzik A., Gough J., Grimmond S., Guetincich S., Hirokawa N., Jackson I.J., Jarvis E.D., Kanai A., Kawaji H., Kawasawa Y., Kedzierski R.M., King B.L., Konagaya A., Kurochkin I.V., Lee Y., Lenhard B., Lyons P.A., Maglott D.R., Maltais L., Marchionni L., McKenzie L., Miki H., Wilming I.G., Wynshaw-Boris A., Yanagisawa M., Yang I., Yang L., Yuan Z., Zavalan M., Zhu Y., Zimmer A., Carninci P., Hayatsu N., Hironaka-Kishikawa T., Konno H., Nakamura M., Sakazume N., Sato K., Shiraki T., Waki K., Kawai J., Aizawa K., Arakawa T., Fukuda S., Hara A., Hashizume W., Imotani K., Ishii Y., Itoh M., Kagawa I., Miyazaki A., Sakai K., Sasaki D., Shibata K., Shingawa A., Yasunishi A., Yoshino M., Waterston R., Lander E.S., Rogers J., Birney E., Hayashizaki Y.;
RT "Analysis of the mouse transcriptome based on functional annotation of 60,770 full-length cDNAs."
RL Nature 420:563-573(2002).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=FVB/N; TISSUE=Mammary gland;
RX MEDLINE=22388257; PubMed=12477932;
RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G., Klausner R.D., Collins P.S., Wagner L., Shenmen C.M., Schuler G.D., Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.P., Bhat N.K., Hopkins R.F., Jordan H., Moore F., Max S.I., Wang J., Hong F., Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L., Stapleton M., Soares M.B., Donaldo M.F., Casavant T.L., Scheetz T.E.,
```

RA Brownstein M.J., Uesdin T.B., Toshiyuki S., Carninci P., Prange C.,
RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullaly S.J.,
RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gouaratne P.H.,
RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Huiyk S.W.,
RA Villalón D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA Fahney J., Helton E., Ketterman M., Madan A., Rodrigues S., Sanchez A.,
RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,
RA Butterfield Y.S.N., Krzywinski M.I., Skalska U., Smalilus D.E.,
RA Schnur A., Schein J.E., Jones S.J.M., Marra M.A.,
RT human and mouse cDNA sequences."
RT generation and initial analysis of more than 15,000 full-length
RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
CC -!- FUNCTION: eIF-2 functions in the early steps of protein synthesis
CC by forming a ternary complex with GTP and initiator tRNA. This
CC complex binds to a 40S ribosomal subunit, followed by mRNA binding
CC to form a 43S preinitiation complex. Junction of the 60S ribosomal
CC subunit to form the 80S initiation complex is preceded by
CC hydrolysis of the GTP bound to eIF-2 and release of an eIF-2-GDP
CC binary complex. In order for eIF-2 to recycle and catalyze another
CC round of initiation, the GDP bound to eIF-2 must exchange with GTP
CC by way of a reaction catalyzed by eIF-2b (By similarity).
CC -!- SUBUNIT: Heterotrimer composed of an alpha, a beta and a gamma
CC chain (By similarity).
CC -!- SIMILARITY: Belongs to the eIF-2-beta / eIF-5 family.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AK011503; BAB27663.1; -
DR EMBL; AK012817; BAB28490.2; ALT_INIT.
DR EMBL; BC003848; AAB03848.1; -
DR WGD; MGI:1914454; Eif2s2.
DR InterPro; IPR002735; eIF5_eIF2B.
DR Pfam; PF01873; eIF5_eIF2B; 1.
DR ProDom; PD004078; eIF5_eIF2B; 1.
DR SMART; SM00653; eIF2B 5; 1.
DR SWART; SM00653; eIF2B 5; 1.
KW Initiation factor; Protein biosynthesis; Zinc-finger.
FT DOMAIN 14 21 POLY-LYS (BASIC).
FT DOMAIN 79 87 POLY-LYS (BASIC).
FT DOMAIN 124 129 POLY-LYS (BASIC).
FT ZN FING 279 303 C4-TYPE (POTENTIAL).
SQ SEQUENCE 331 AA; 38092 MW; E91BEBBF330B7A5C CRC64;

Query Match 1.3%; Score 7; DB 1; Length 331;
Best Local Similarity 100.0%; E-val No. 80;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 87 EPAEPED 93
Db 109 EPAEPED 115

RESULT 52
ID IF2B_RABIT STANDARD; PRT; 333 AA.
AC F41035;
DT 01-FEB-1995 (Rel. 31, Created)
DT 01-FEB-1995 (Rel. 31, Last sequence update)
DT 15-MAR-2004 (Rel. 43, Last annotation update)
DE Eukaryotic translation initiation factor 2 subunit 2 (Eukaryotic
DE translation initiation factor 2 beta subunit) (eIF-2-beta).
GN EIF2B2 OR EIF2B.
OS Oryctolagus cuniculus (Rabbit).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Lagomorpha; Leporidae; Oryctolagus.
OX NCBI_TaxID=9986;

RN SEQUENCE FROM N.A.
RC STRAIN=New Zealand white.
RX MEDLINE=94032484; PubMed=8218412;
RA Price N.T., Hall L., Proud C.G.;
RT "Cloning of cDNA for the beta-subunit of rabbit translation initiation
RL factor-2 using PCR.";
RN Biochim. Biophys. Acta 1216:170-172(1993).
RP [2]
RP SEQUENCE OF 25-39; 122-136 AND 195-218.
RC TISSUE=Reticulocytes;
RX MEDLINE=92002150; PubMed=1911855;
RA Bommer U.-A., Kraft R., Kurzchalia T.V., Price N.T., Proud C.G.;
RT "Amino acid sequence analysis of the beta- and gamma-subunits of
RL eukaryotic initiation factor eIF-2. Identification of regions
RN interacting with GTP.";
RP Biochim. Biophys. Acta 1079:308-315(1991).
RP [3]
RP CHARACTERIZATION.
RC TISSUE=Reticulocytes;
RX MEDLINE=87240215; PubMed=3592677;
RA Schafer M.P., Fairwell T., Parker D.S., Knight M., Anderson W.F.,
RT "The purification and characterization of subunits alpha, beta, and
RL gamma from the rabbit reticulocyte eukaryotic initiation factor 2.";
CC Arch. Biochem. Biophys. 255:337-346(1987).
CC -!- FUNCTION: eIF-2 functions in the early steps of protein synthesis
CC by forming a ternary complex with GTP and initiator tRNA. This
CC complex binds to a 40S ribosomal subunit, followed by mRNA binding
CC to form a 43S preinitiation complex. Junction of the 60S ribosomal
CC subunit to form the 80S initiation complex is preceded by
CC hydrolysis of the GTP bound to eIF-2 and release of an eIF-2-GDP
CC binary complex. In order for eIF-2 to recycle and catalyze another
CC round of initiation, the GDP bound to eIF-2 must exchange with GTP
CC by way of a reaction catalyzed by eIF-2b.
CC -!- SUBUNIT: Heterotrimer composed of an alpha, a beta and a gamma
CC chain.
CC -!- PTM: The N-terminus is blocked.
CC -!- SIMILARITY: Belongs to the eIF-2-beta / eIF-5 family.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; X73836; CAA52058.1; -
DR PIR; JC2329; JC2329.
DR PIR; S17871; S17871.
DR InterPro; IPR002735; eIF5_eIF2B.
DR Pfam; PF01873; eIF5_eIF2B; 1.
DR ProDom; PD004078; eIF5_eIF2B; 1.
DR SMART; SM00653; eIF2B 5; 1.
KW Initiation factor; Protein biosynthesis; Zinc-finger.
FT DOMAIN 14 21 POLY-LYS (BASIC).
FT DOMAIN 79 87 POLY-LYS (BASIC).
FT DOMAIN 124 129 POLY-LYS (BASIC).
FT ZN FING 281 305 C4-TYPE (POTENTIAL).
SQ SEQUENCE 333 AA; 38326 MW; 3FFDBAB92DFC1465 CRC64;

Query Match 1.3%; Score 7; DB 1; Length 333;
Best Local Similarity 100.0%; E-val No. 81;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 87 EPAEPED 93
Db 109 EPAEPED 115

RESULT 53
SYFA_CLOAB

ID SYFA CLOAB STANDARD; PRT; 339 AA.
AC Q97GK9;
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Phenylalanyl-tRNA synthetase alpha chain (EC 6.1.1.20)
DE (Phenylalanine-tRNA ligase alpha chain) (PHERS).
GN PHERS OR CAC2357.
OS Clostridium acetobutylicum.
OC Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;
OC Clostridium.
OX NCBI_TaxID=1488;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 824 / DSM 792 / VKM B-1787;
RX MEDLINE=2135925; PubMed=11466286;
RA Noelling J., Bregon G., Omelchenko M.V., Makarova K.S., Zeng Q.,
RA Gibson R., Lee H.M., Dubois J., Qiu D., Hitti J., Wolf Y.I.,
RA Tatusov R.L., Sabathe F., Doucette-Stamm L., Soucaille P., Daly M.J.,
RA Bennett G.N., Koonin E.V., Smith D.R.;
RT "Genome sequence and comparative analysis of the solvent-producing
bacterium Clostridium acetobutylicum.";
RL J. Bacteriol. 183:4823-4838 (2001).
CC -!- CATALYTIC ACTIVITY: ATP + L-phenylalanine + tRNA(Phe) = AMP +
CC diphosphate + L-phenylalanyl-tRNA(Phe).
CC -!- COPACTOR: Binds 2 magnesium ions per tetramer (By similarity).
CC -!- SUBUNIT: Tetramer of two alpha and two beta chains (By
CC similarity).
CC -!- SUBCELLULAR LOCATION: Cytoplasmic.
CC -!- SIMILARITY: Belongs to class-II aminoacyl-tRNA synthetase family.
CC Phe-tRNA synthetase alpha chain subfamily 1.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).

CC EMBL; AE007736; AAK80313.1; -
CC PIR; F97190; F97190.
CC HAMAP; MF_00281; -; 1.
CC InterPro; IPR004188; Phe-tRNA_synth_N.
CC InterPro; IPR004529; PheS.
CC InterPro; IPR002319; tRNA-synth_2d.
CC InterPro; IPR006195; tRNA-synth_N; I.
CC Pfam; PF02912; Phe-tRNA-synth_N; I.
CC Pfam; PF01409; tRNA-synth_2d; I.
CC TIGRFAMs; TIGR00468; pheS; I.
CC PROSITE; PSS0862; AA-tRNA_LIGASE II; 1.
KW Aminoacyl-tRNA synthetase; Protein biosynthesis; Ligase; ATP-binding;
KW Metal-binding; Magnesium; Complete proteome.
FT METAL 254 254 MAGNESIUM (BY SIMILARITY).
SQ SEQUENCE 339 AA; 38508 MW; 229E76449BBE10918 CRC64;

Query Match 1.3%; Score 7; DB 1; Length 339;
Best Local Similarity 100.0%; Pred. No. 82;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 222 RSKLES 228
Db 65 RSKLES 71
|||||
RESULT 54
IAP_ECOLI
ID IAP_ECOLI STANDARD; PRT; 345 AA.
AC P10423;
DT 01-MAR-1989 (Rel. 10, Created)
DT 01-MAR-1989 (Rel. 10, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Alkaline phosphatase isozyme conversion protein precursor

(EC 3.4.11.1).
DE IAP OR B2753.
GN Escherichia coli.
OS Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Escherichia.
OX NCBI_TaxID=562;
RN [1]
RP SEQUENCE FROM N.A.
RC MEDLINE=88058748; PubMed=3316184;
RX Ishino Y., Shinagawa H., Makino K., Anemura M., Nakata A.;
RT "Nucleotide sequence of the iap gene, responsible for alkaline
RT phosphatase isozyme conversion in Escherichia coli, and
RT identification of the gene product.";
RL J. Bacteriol. 169:5429-5433 (1987).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=K12 / MGL1655;
RX MEDLINE=97426617; PubMed=9278503;
RA Blattner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,
RA Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,
RA Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,
RA Mau B., Shao Y.;
RT "The complete genome sequence of Escherichia coli K-12.";
RL Science 277:1453-1474 (1997).
RN [3]
RP SEQUENCE OF 1-111 FROM N.A.
RC STRAIN=K12;
RX MEDLINE=92268080; PubMed=1316900;
RA Leyh T.S., Vogt T.F., Suo Y.;
RT "The DNA sequence of the sulfate activation locus from Escherichia
RT coli K-12.";
RL J. Biol. Chem. 267:10405-10410 (1992).
CC -!- FUNCTION: This protein, presumably an aminopeptidase, mediates the
CC conversion of E.coli alkaline phosphatase isozyme 1, to isozymes 2
CC and 3 by removing, one by one, the two amino-terminal arginine
CC residues.
CC -!- SIMILARITY: Belongs to peptidase family M28C.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).

CC EMBL; M18270; AAK24005.1; -
CC EMBL; U29579; AAK69263.1; -
CC EMBL; AE000359; AAC75795.1; -
CC EMBL; M74586; AAA23644.1; -
CC PIR; A28382; KZEC.
CC MEROPS; M28.005; -
CC EcoGene; EG10488; iap.
CC InterPro; IPR007484; Peptidase_M28.
CC Pfam; PF04389; Peptidase_M28; I.
KW Hydrolase; Aminopeptidase; Signal; Complete proteome.
FT SIGNAL 1 24 PROBABLE.
FT CHAIN 25 345 ALKALINE PHOSPHATASE ISOZYME CONVERSION
FT PROTEIN.
SQ SEQUENCE 345 AA; 37920 MW; 3DC433FB569E912D CRC64;

Query Match 1.3%; Score 7; DB 1; Length 345;
Best Local Similarity 100.0%; Pred. No. 83;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 288 ELAERLK 294
Db 153 ELAERLK 159
|||||
RESULT 55
RF1_TREPA
ID RF1_TREPA STANDARD; PRT; 351 AA.

```

AC 083090;
DT 15-DEC-1998 (Rel. 37, Created)
DT 15-DEC-1998 (Rel. 37, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Peptide chain release factor 1 (RP-1).
GN PRFA OR TP0051.
OS Treponema pallidum.
OC Bacteria; Spirochaetes; Spirochaetales; Spirochaetaceae; Treponema.
OX NCBI_TaxID=160;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C-125 / JCM 9153;
RX MEDLINE=20512582; PubMed=11058132;
RA Takami H., Nakasone K., Takaki Y., Maeno G., Sasaki R., Masui N.,
RA Fuji P., Hirama C., Nakamura Y., Ogasawara N., Kuhara S.,
RA Horikoshi K.;
RT "Complete genome sequence of the alkaliphilic bacterium Bacillus
RT halodurans and genomic sequence comparison with Bacillus subtilis.";
RL Nucleic Acids Res. 28:4317-4331(2000).
CC -!- FUNCTION: Peptide chain release factor 1 directs the termination
CC of translation in response to the peptide chain termination codons
CC UAG and UAA (By similarity).
CC -!- SUBCELLULAR LOCATION: Cytoplasmic.
CC -!- SIMILARITY: Belongs to the prokaryotic/mitochondrial release
CC factor family.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AP001519; BAB07494.1; -
DR PIR; G84121; G84121.
DR HAMAP; MF_00093; -; 1.
DR InterPro; IPR005139; PCRF.
DR InterPro; IPR000352; Pep_rel_factor_1.
DR InterPro; IPR004373; PrfA.
DR Pfam; PF03462; PCRF; 1.
DR Pfam; PF00472; RF-1; 1.
DR TIGRFAMs; TIGR00019; prfA; 1.
DR PROSITE; PS00745; RF_PROK_1; 1.
DR Protein biosynthesis; Complete proteome.
KW SEQUENCE 356 AA; 40597 MW; 34342F09CBBAA8 CRC64;
SQ
Query Match 1.3%; Score 7; DB 1; Length 356;
Best Local Similarity 100.0%; Pred. No. 86;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 358 ELMKQOE 364
Db 349 ELMKQOE 355
|||||
RESULT 57
MKK2 DROME
ID MKK2 DROME STANDARD; PRT; 359 AA.
AC P49071; Q9W480;
DT 01-FEB-1996 (Rel. 33, Created)
DT 01-FEB-1996 (Rel. 33, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DB MAP kinase-activated protein kinase 2 (EC 2.7.1.-) (MAPKAPK-2)
DE protein kinase 2 (MAPKAP kinase 2) (MAPKAPK-2).
GN MAPK-AK2 OR CG3086.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=BIC-D(R26); TISSUE=Ovary;
RX MEDLINE=96011635; PubMed=7590268;
RA Laroche S., Suter B.;
RT "The Drosophila melanogaster homolog of the mammalian MAPK-activated
RT protein kinase-2 (MAPKAPK-2) lacks a proline-rich N-terminus.";
RL Gene 163:209-214(1995).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=Berkeley;

```

```

RX MEDLINE=20196006; PubMed=10731132;
RA Adams M.D., Ceiniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.M.,
RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA Brandon R.C., Rogers Y.-H.C., Blazej R.G., Chape M., Pfeiffer B.D.,
RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA Abril J.P., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
RA Bailew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA Borokova D., Botchan M.R., Bouck J., Brokstein P., Brotter P.,
RA Burtis K.C., Busan D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA de Pablo B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA Durbin K., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,
RA Flossler C., Gabriellian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA Harris N.L., Harvey D.A., Heiman T.J., Hernandez J.R., Houck J.,
RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA Laoko P., Lei Y., Leviscky A.A., Li J.H., Li Z., Liang Y., Lin X.,
RA Liu X., Mattel B., McIntosh T.C., McLeod M.P., McPherson D.,
RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Paclib J.M.,
RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA Shue B.C., Sidel-Kiamis I., Simpson M., Skupski M.P., Smith T.,
RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun B.,
RA Svirskas R., Tector C., Turner R., Venter Z., Wang A.H., Wang X.,
RA Wang Z.-Y., Wassarman D.A., Weinstein G.M., Weissbach J.,
RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA Zhang X.H., Zhong F.N., Zhong W., Zhou X., Zhu X., Smith H.O.,
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RA "The genome sequence of Drosophila melanogaster."
RT Science 287:2185-2195(2000).
CC -!- FUNCTION: ITS PHYSIOLOGICAL SUBSTRATE SEEMS TO BE THE SMALL HEAT
CC SHOCK PROTEIN (HSP70/HSP25) (BY SIMILARITY).
CC -!- PTM: Phosphorylated and activated by MAP kinase (By similarity).
CC -!- SIMILARITY: Belongs to the Ser/Thr family of protein kinases.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; U20757; AAA86885.1; -
DR EMBL; AE003435; AAG22408.1; -
DR PIR; JC4297; JCA297.
DR HSP; Q63450; 1A06.
DR FlyBase; FBGT0013987; MAPK-AK2.
DR InterPro; IPR000719; Prot_kinase.
DR InterPro; IPR008271; Ser_thr_pkin_AS.
DR InterPro; IPR002290; Ser_thr_pkinase.
DR InterPro; IPR001245; Tyr_pkinase.
DR Pfam; PF00069; pkinase; 1.
DR PRINTS; PR00109; TYRKINASE.
DR ProDom; PD000001; Prot_kinase; 1.
DR SMART; SM00220; S_TKC; 1.
DR PROSITE; PS00107; PROTEIN KINASE ATP; 1.
DR PROSITE; PS00108; PROTEIN KINASE ST; 1.
DR PROSITE; PSS0011; PROTEIN KINASE DOM; 1.
DR TRANSFERASE; Serine/threonine-protein kinase; ATP-binding;
KW Phosphorylation.
FT DOMAIN 20 281 PROTEIN KINASE.
FT NP_BIND 26 34 ATP (BY SIMILARITY).
FT BINDING 49 49 ATP (BY SIMILARITY).
CC -----
FT ACT_SITE 142 142 BY SIMILARITY.
SQ SEQUENCE 359 AA; 41401 MW; 69CSF9A94D511EC CRC64;
Query Match 1.3%; Score 7; DB 1; Length 359;
Best Local Similarity 100.0%; Pred. No. 86;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 331 MLKEAE 337
DB 297 MLKEAE 303
|||||
RESULT 58
YHNG YEAST STANDARD; PRT; 374 AA.
AC P38797;
DT 01-FEB-1995 (Rel. 31, Last sequence update)
DT 01-FEB-1995 (Rel. 31, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Hypothetical 41.2 kDa protein in ERG7-NMD2 intergenic region.
GN YHR076W.
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
OX NCBI_TaxID=4932;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=S288C / AB972;
RX MEDLINE=94378003; PubMed=8091223;
RA Johnston M., Andrews S., Brinkman R., Cooper J., Ding H., Dover J.,
RA Du Z., Favella A., Fulton L., Gattung S., Geisel C., Kirsten J.,
RA Kucaba T., Hillier L.W., Jier M., Johnston L., Langston Y.,
RA Latreille P., Louis E.J., Macri C., Mardis E., Menezes S., Mouser L.,
RA Nghan M., Rifkin L., Riles L., St Peter H., Trevaskis E., Vaughan K.,
RA Vignati D., Wilcox L., Wohldman P., Waterston R., Wilson R.,
RA Vaudin M.;
RT "Complete nucleotide sequence of Saccharomyces cerevisiae chromosome
RT VIII."
RL Science 265:2077-2082(1994).
CC -!- SIMILARITY: Contains 1 PP2C-like domain.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; U10556; AAB68888.1; -
DR PIR; S46810; S46810.
DR GerMOnline; 139393; -
DR SGD; S0001118; YHR076W.
DR GO; GO:0005739; C:mitochondrion; IDA.
DR GO; GO:0015071; P:protein phosphatase type 2C activity; IDA.
DR InterPro; IPR001932; PP2C-SIG; 1.
DR SMART; SM00331; PP2C_SIG; 1.
DR SMART; SM00332; PP2C; 1.
DR TRANSFERASE; Hypothetical protein; Transmembrane.
FT TRANSMEM 19 39 POTENTIAL.
FT DOMAIN 117 374 PP2C-LIKE.
SQ SEQUENCE 374 AA; 41190 MW; 74E7F084ABD157FE CRC64;
Query Match 1.3%; Score 7; DB 1; Length 374;
Best Local Similarity 100.0%; Pred. No. 89;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 330 EMLKEAE 336
DB 258 EMLKEAE 264
|||||
RESULT 59
```

```

OC Eutotiales; Trichocomaceae; mitosporic Trichocomaceae; Aspergillus.
OX NCBI_TaxID=5059;
[1]
RN
RP SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
RX MEDLINE=94086131; PubMed=8262652;
RA Ramesh M.V., Sirakova T., Koiattukudy P.E.;
RT "isolation, characterization, and cloning of cDNA and the gene for an
RL elastolytic serine proteinase from Aspergillus flavus.";
RL Infect. Immun. 62:79-85(1994).
CC -!- FUNCTION: Elastolytic enzyme probably acting as a virulence factor
CC in invasive aspergillosis.
CC -!- CATALYTIC ACTIVITY: Hydrolysis of proteins with broad specificity,
CC and of BZ-Arg-ORF > Ac-Tyr-ORF. Does not hydrolyze peptide amides.
CC -!- SUBCELLULAR LOCATION: Secreted.
CC -!- SIMILARITY: Belongs to peptidase family S8.
-----
CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
-----
CC EMBL; L08473; AAA32691.1; --
CC EMBL; S67840; AAB29384.1; --
CC HSP; P06873; 2PRK.
CC MEROPS; S08.053; --
CC InterPro; IPR000209; Peptidase S8.
CC InterPro; IPR009020; Protease inhib.
CC Pfam; PF00082; Peptidase S8; 1.
CC PRINTS; PR00723; SUBTILISIN.
CC PROSITE; PS00136; SUBTILASE ASP; FALSE_NEG.
CC PROSITE; PS00137; SUBTILASE_HIS; 1.
CC PROSITE; PS00138; SUBTILASE_SER; 1.
CC Hydrolase; Serine protease; Glycoprotein; Zymogen; Signal.
KW SIGNAL 1 21
FT PROPEP 22 121
FT CHAIN 122 403
FT ACT_SITE 162 162
FT ACT_SITE 193 193
FT ACT_SITE 349 349
FT CARBOHYD 253 253
FT CARBOHYD 309 309
SQ SEQUENCE 403 AA; 42768 MW; D9B86FFBA26107C6 CRC64;
Query Match 1.3%; Score 7; DB 1; Length 403;
Best Local Similarity 100.0%; Pred.No. 95;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 129 IRQSDEV 135
Db 103 IRQSDEV 109
|||||
-----
RESULT 61
AROA THETN STANDARD; PRT; 423 AA.
AC AROA11;
AC QSRB11;
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE 3-phosphoshikimate 1-carboxyvinyltransferase (EC 2.5.1.19) (5-
DE enolpyruvylshikimate-3-phosphate synthase) (BPS synthase) (BPSPS).
OS AROA OR AROA2 OR TFE1015.
OS Thermoanaerobacter tengcongensis.
OC Bacteria; Firmicutes; Clostridia; Thermoanaerobacteriales;
OC Thermoanaerobacteriaceae;
OC Thermoanaerobacter.
OX NCBI_TaxID=119072;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN=MB4 / JCM 11007;
RX MEDLINE=21992816; PubMed=11997336;

```

RA Bao Q., Tian Y., Li W., Xu Z., Xuan Z., Hu S., Dong W., Yang J.,
RA Chen Y., Xue Y., Lai X., Huang L., Dong X., Ma Y., Ling L.,
RA Tan H., Chen R., Wang J., Yu J., Yang H.,
RT "A complete sequence of T. tengcongensis genome.";
RL Genome Res. 12:689-700(2002).
CC -1- CATALYTIC ACTIVITY: Phosphoenolpyruvate + 3-phosphohikimate =
CC phosphate + 5-O-(1-carboxyvinyl)-3-phosphohikimate.
CC -1- PATHWAY: Aromatic amino acids biosynthesis; shikimate pathway;
CC sixth step.
CC -1- SUBUNIT: Monomer (By similarity).
CC -1- SUBCELLULAR LOCATION: Cytoplasmic (Probable).
CC -1- SIMILARITY: Belongs to the EPSP synthase family.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC
DR EMBL; AE013066; AAM24270.1; -
DR HAMAP; MF_00210; -; 1.
DR InterPro; IPR006264; AroA.
DR Pfam; PF00275; EPSP_synthase; 1.
DR ProDom; PD001867; EPSP_synthase; 1.
DR TIGRFAMs; TIGR01356; aroA; 1.
DR PROSITE; PS00104; EPSP_SYNTHASE_1; 1.
DR PROSITE; PS00885; EPSP_SYNTHASE_2; 1.
KW Aromatic amino acid biosynthesis; Transferrase; Complete proteome.
SQ SEQUENCE 423 AA; 46069 MW; C8BECDC1902DF51 CRC64;

Query Match 1.3%; Score 7; DB 1; Length 423;
Best Local Similarity 100.0%; Pred. No. 1e+02; 0; Indels 0; Gaps 0;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 385 LSKSSEV 391
DB 79 LSKSSEV 85
|||||||

RESULT 62
Y12A_ECOLI
ID Y12A_ECOLI STANDARD; PRT; 424 AA.
AC P51026; Q47535;
DT 01-OCT-1996 (Rel. 34; Created)
DT 15-JUL-1999 (Rel. 38; Last sequence update)
DT 16-OCT-2001 (Rel. 40; Last annotation update)
DE Insertion element IS2A hypothetical 48.2 kDa protein.
OS Escherichia coli.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Escherichia.
OX NCBI_TaxID=562;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=K12;
RA Nashimoto H., Saito N.;
RL Submitted (JUN-1996) to the EMBL/GenBank/DBJ databases.
CC -1- SIMILARITY: Belongs to the transposase family 8.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC
DR EMBL; D85613; BAA12836.1; -
DR FIR; S78610; S78610.
DR InterPro; IPR001584; Rve.
DR InterPro; IPR002514; Transposase_8.

DR Pfam; PF00665; rve; 1.
DR Pfam; PF01527; Transposase_8; 1.
KW Transposable element; Hypothetical protein.
SQ SEQUENCE 424 AA; 48153 MW; A791CE85D05A329 CRC64;

Query Match 1.3%; Score 7; DB 1; Length 424;
Best Local Similarity 100.0%; Pred. No. 1e+02; 0; Indels 0; Gaps 0;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 346 LKKEAVE 352
DB 111 LKKEAVE 117
|||||||

RESULT 63
FIBG_PETMA
ID FIBG_PETMA STANDARD; PRT; 432 AA.
AC P04115;
DT 01-NOV-1986 (Rel. 03; Created)
DT 01-NOV-1986 (Rel. 03; Last sequence update)
DT 10-OCT-2003 (Rel. 42; Last annotation update)
DE Fibrinogen gamma chain precursor.
OS Petromyzon marinus (Sea lamprey).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Hyperoartia;
OC Petromyzontiformes; Petromyzontidae; Petromyzon.
OX NCBI_TaxID=7757;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=8519776; PubMed=2581603;
RA Strong D.D., Moore M., Cottrell B.A., Bohonus V.L., Pontes M.,
RA Evans B., Riley K., Doellittle R.F.;
RT "Lamprey fibrinogen gamma chain: cloning, cDNA sequencing, and
RT general characterization.";
RL Biochemistry 24:92-101(1985).
CC -1- FUNCTION: Fibrinogen has a double function: yielding monomers that
CC polymerize into fibrin and acting as a cofactor in platelet
CC aggregation.
CC -1- SUBUNIT: HEXAMER CONTAINING 2 SETS OF 3 NONIDENTICAL CHAINS
CC (ALPHA, BETA AND GAMMA), LINKED TO EACH OTHER BY DISULFIDE BONDS.
CC -1- PTM: Conversion of fibrinogen to fibrin is triggered by thrombin,
CC which cleaves fibrinopeptides A and B from alpha and beta chains,
CC and thus exposes the N-terminal polymerization sites responsible
CC for the formation of the soft clot. The soft clot is converted
CC into the hard clot by factor XIIIa which catalyzes the epsilon-
CC (gamma-glutamyl)lysine cross-linking between gamma chains
CC (stronger) and between alpha chains (weaker) of different
CC monomers.
CC -1- SIMILARITY: Contains 1 fibrinogen C-terminal domain.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC
DR EMBL; K03049; AAA49262.1; -
DR FIR; A03129; FGLMGS.
DR PDB; 1LWU; 23-AUG-02.
DR PDB; 1N73; 07-JAN-03.
DR InterPro; IPR002181; Fibrinogen_C.
DR Pfam; PF00147; fibrinogen_C; 1.
DR SMART; SM00186; FBG; 1.
DR PROSITE; PS00514; FIBRIN_AG_C_DOMAIN; 1.
KW Blood coagulation; Plasma; Glycoprotein; Signal; 3D-structure.
FT SIGNAL 1 24 BY SIMILARITY
FT CHAIN 25 432 FIBRINOGEN GAMMA CHAIN
FT DISULFID 32 32 INTERCHAIN (WITH GAMMA CHAIN)
FT DISULFID 42 42 INTERCHAIN (WITH BETA CHAIN)
FT DISULFID 46 46 INTERCHAIN (WITH ALPHA CHAIN)

FT DISULFID 158 158 (BY SIMILARITY).
FT INTERCHAIN (WITH BETA CHAIN)
FT (BY SIMILARITY).
FT INTERCHAIN (WITH GAMMA CHAIN)
FT (BY SIMILARITY).
FT DISULFID 178 207
FT DISULFID 348 361
FT CARBOHYD 227 227
SQ SEQUENCE 432 AA; 49203 MW; B503979B296DFB24 CRC64;
Query Match 1.3%; Score 7; DB 1; Length 432;
Best Local Similarity 100.0%; Pred. No. le+02; Indels 0; Gaps 0;
Matches 7; Conservative 0; Mismatches 0;
QY 316 LQQQLVD 322
DB 147 LQQQLVD 153
RESULT 64
SDHD_FUSNN
ID SDHD_FUSNN STANDARD; PRT; 441 AA.
AC Q8RFK6;
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Probable D-serine dehydratase (EC 4.3.1.18) (D-serine deaminase)
DE (DSB).
GN DSDA OR FN0553.
OS Fusobacterium nucleatum (subsp. nucleatum).
OC Bacteria; Fusobacteria; Fusobacteriales; Fusobacteriaceae;
OC Fusobacterium
OX NCBI_TaxID=76956;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 25586;
RX MEDLINE=21886394; PubMed=11889109;
RA Kapatal V., Anderson I., Ivanova N., Reznik G., Los T., Lykidis A.,
RA Bhattacharyya A., Bartman A., Gardner W., Grechkin G., Zhu L.,
RA Vasileva O., Chu L., Kogan Y., Chaga O., Goltsman E., Bernal A.,
RA Larsen N., D'Souza M., Walunas T., Fusch G., Hasekorn R.,
RA Fongstein M., Kyriakides N., Overbeek R.,
RT "Genome sequence and analysis of the oral bacterium Fusobacterium
nucleatum strain ATCC 25586.";
RL J. Bacteriol. 184:2005-2018(2002).
CC -!- CATALYTIC ACTIVITY: D-serine = pyruvate + NH(3).
CC -!- COFACTOR: Pyridoxal phosphate (By similarity).
CC -!- SIMILARITY: Belongs to the serine/threonine dehydratase family.
DSDA subfamily.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@sib-sib.ch).

DR EMBL; AE010566; AAL94749.1;
DR HAMAP; MF_01030; -; 1
DR InterPro; IPR001926; S6 enzyme_beta.
DR InterPro; IPR000634; S/T_dehydrtse_BS.
DR Pfam; PF00291; PALP; 1.
DR PROSITE; PS00165; DEHYDRATASE_SRR_THR; FALSE NEG.
KW Lyase; Pyridoxal phosphate; Complete proteome.
FT BINDING 115 115 PYRIDOXAL PHOSPHATE (BY SIMILARITY).
SQ SEQUENCE 441 AA; 49954 MW; E2058AF21B4851E2 CRC64;
Query Match 1.3%; Score 7; DB 1; Length 441;
Best Local Similarity 100.0%; Pred. No. le+02; Indels 0; Gaps 0;
Matches 7; Conservative 0; Mismatches 0;
QY 332 LKEAER 338

DB 46 LKEAER 52
RESULT 65
GAS8_HUMAN
ID GAS8_HUMAN STANDARD; PRT; 478 AA.
AC O95995;
DT 10-OCT-2003 (Rel. 42, Created)
DT 10-OCT-2003 (Rel. 42, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Growth-arrest-specific protein 8 (Growth arrest-specific 11).
DE GAS8 OR GAS11.
GN Homo sapiens (Human)
OS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=99009317; PubMed=9790751;
RA Whitmore S.A., Seltastian C., Crawford J., Lower K.M., McCallum B.,
RA Seshadri R., Cornelisse C.J., Moerland E.W., Cleton-Jansen A.-M.,
RA Tipping A.J., Mathew C.G., Savino M., Savoia A., Verlander P.,
RA Auerbach A.D., Van Berkel C., Pronk J.C., Doggett N.A., Calien D.F.;
RT "Characterization and screening for mutations of the growth arrest-
RT specific 11 (GAS11) and C16orf3 genes at 16q24.3 in breast cancer.";
RL Genomics 52:325-331(1998).
RN [2]
MP MICROTUBULE-BINDING.
RP MEDLINE=20564307; PubMed=10969087;
RX Hill K.L., Hutchings N.R., Grandgenett P.M., Donelson J.E.;
RT "T lymphocyte-triggering factor of African trypanosomes is associated
RT with the flagellar fraction of the cytoskeleton and represents a new
RT family of proteins that are present in several divergent
RT eukaryotes.";
RL J. Biol. Chem. 275:39369-39378(2000).
RN [3]
DISEASE.
RX MEDLINE=21839108; PubMed=11751847;
RA Yeh S.-D., Chen Y.-J., Chang A.C.Y., Ray R., She B.-R., Lee W.-S.,
RA Chiang H.-S., Cohen S.N., Lin-Chao S.;
RT "Isolation and properties of Gas8, a growth arrest-specific gene
RT regulated during male gametogenesis to produce a protein associated
RT with the sperm motility apparatus.";
RL J. Biol. Chem. 277:6311-6317(2002).
CC -!- FUNCTION: Cytoskeletal linker that probably plays a role in the
CC spermatozoa motility (By similarity).
CC -!- SUBCELLULAR LOCATION: Cytoplasmic; associates with microtubules.
CC In spermatozoa, it localizes to the flagellum (By similarity).
CC -!- TISSUE SPECIFICITY: Expressed in the heart, skeletal muscle,
CC pancreas and liver. Weakly or not expressed in brain, placenta,
CC lung and kidney.
CC -!- DISEASE: Defects in GAS8 may be a cause of infertility in male
CC lacking gametocytes.
CC -!- SIMILARITY: Belongs to the GAS8 family.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@sib-sib.ch).

DR EMBL; AF050079; AAC69519.1;
DR EMBL; AF050078; AAC69518.1; JOINED.
DR EMBL; AF050068; AAC69518.1; JOINED.
DR EMBL; AF050065; AAC69518.1; JOINED.
DR EMBL; AF050070; AAC69518.1; JOINED.
DR EMBL; AF050071; AAC69518.1; JOINED.
DR EMBL; AF050072; AAC69518.1; JOINED.
DR EMBL; AF050073; AAC69518.1; JOINED.
DR EMBL; AF050074; AAC69518.1; JOINED.

RA Unayam L.A., White O., Salzberg S.L., Lewis M.R., Radune D., Holtzapfel E., Khouri H., Wolf A.M., Utterback T.R., Hansen C.L., McDonald L.A., Feldblyum T.V., Anguoli S., Dickinson T., Hickey E.K., Holt I.E., Lofcus B.J., Yang P., Smith H.O., Venter J.C., Dougherty B.A., Morrison D.A., Hollingshead S.K., Fraser C.M., "Complete genome sequence of a virulent isolate of Streptococcus pneumoniae"; Science 293:498-506(2001).

RL [2]

RN SEQUENCE OF 281-433 FROM N.A.

RP STRAIN=ATCC 6323;

RC MEDLINE=98125733; PubMed=946257;

RX Coffey T.J., Enright M.C., Daniels M., Morona J.K., Morona R., Ryniewicz W., Paton J.C., Spratt B.G.; "Recombinational exchanges at the capsular polysaccharide biosynthetic locus lead to frequent serotype changes among natural isolates of Streptococcus pneumoniae"; Mol. Microbiol. 27:73-83(1998).

RL [3]

RN SEQUENCE OF 281-433 FROM N.A.

RP MEDLINE=98192409; PubMed=9533725;

RX Coffey T.J., Enright M.C., Daniels M., Wilkinson P., Berron S., Fenoll A., Spratt B.G.; "Serotype 19A variants of the Spanish serotype 23F multiresistant clone of Streptococcus pneumoniae"; Microb. Drug Resist. 4:51-55(1998).

RT -1- CATALYTIC ACTIVITY: D-glucose 6-phosphate + NADP(+) = D-gluconolactone + NADPH.

CC -1- LACTONE 6-phosphate + NADPH.

CC -1- PATHWAY: Pentose phosphate pathway: first step

CC -1- SIMILARITY: Belongs to the glucose-6-phosphate dehydrogenase family.

CC

CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

CC

DR ENBL; A8007424; AAK75348.1; -

DR ENBL; Z99802; CAB16927.1; -

DR ENBL; Z99865; CAB16990.1; -

DR PIR; C95144; C95144.

DR HSSP; P11411; 1DPG.

DR TIGR; SPI243; -

DR InterPro; IPR001282; G6PD.

DR Pfam; PF00479; G6PD; 1.

DR Pfam; PF02781; G6PD; 1.

DR PRINTS; PR00079; G6PDHGNASE.

DR PRODOM; PD001129; G6PD; 1.

DR TIGRFAMS; TIGR00871; zwf; 1.

DR PROSITE; PS00069; G6P_DEHYDROGENASE; 1.

KW Oxidoreductase; NADP; Glucose metabolism; Complete proteome.

FT ACT_SITE 181 181 BY SIMILARITY.

FT CONFLICT 348 348 V -> I (IN REF. 2 AND 3).

FT SEQUENCE 495 AA; 56790 MW; 990393EC8929FFAF CRC64;

QY

Db 48 QSGALRD 54

228 QSGALRD 234

Query Match 1.3%; Score 7; DB 1; Length 495;

Best Local Similarity 100.0%; Pred. No. 1.1e+02;

Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

RESULT 68

CN24 HUMAN STANDARD; PRT; 496 AA.

AC Q9NVE0; Q8N3D3;

DT 15-MAR-2004 (Rel. 43, Created)

DT 15-MAR-2004 (Rel. 43, Last sequence update)

DT 15-MAR-2004 (Rel. 43, Last annotation update)

DE Protein C14orf114.

GN C14ORF114.

OS Homo sapiens (Human).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

OX NCBI_TaxID=9606;

RN [1]_TaxID=9606;

RP SEQUENCE FROM N.A.

RC TISSUE=Testis; Carcinoma;

RA Isogai T., Ota T., Hayashi K., Sugiyama T., Otsuki T., Suzuki Y., Nishikawa T., Nagai K., Sugano S., Shiratori A., Sudo H., Wagatsuma M., Hosoi T., Kaku Y., Kondo H., Kondo H., Sugawara M., Takahashi M., Chiba Y., Ishida S., Murakawa K., Ono Y., Takiguchi S., Katanabe S., Kimura K., Murakami K., Ishii S., Kawai Y., Saio K., Yamanoto J., Wakamatsu A., Nakamura Y., Nagahari K., Masuho Y., Ninomiya K., Iwayanagi T.; "NEDO human cDNA sequencing project"; NEDO Submitted (FEB-2000) to the EMBL/GenBank/DBJ databases.

RL [2]

RN SEQUENCE FROM N.A.

RP TISSUE=Melanoma;

RC Blocker H., Boecher M., Brandt P., Mewes H.-W., Weil B., Wiemann S.; "Submitted (JUL-2002) to the EMBL/GenBank/DBJ databases.

RL [3]

RN SEQUENCE FROM N.A.

RP TISSUE=Lung;

RC MEDLINE=22388257; PubMed=12477932;

RX Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G., Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D., Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K., Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F., Stapleton M., Marusina K., Farmer A.A., Rubin G.M., Hong L., Chao L., DeLong M., Bosman M., Bonaldo M.F., Casavant T.L., Scheetz T.E., Brownstein M.J., Ustin T.B., Toshiyuki S., Carninci P., Prange C., Raha S.S., Lomellano N.A., Peters G.J., Abramson R.D., Mullaly S.J., Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H., Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W., Villalón D.K., Muzny K.C., Sodergren E.J., Lu X., Gibbs R.A., Fahy J., Helton E., Kettner M., Madan A., Rodriguez S., Sanchez A., Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C., Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M., Butterfield Y.S.N., Krzywinski M.I., Skalska U., Smailus D.E., Scherch A., Schein J.E., Jones S.J.M., Marra M.A.; "Generation and initial analysis of more than 15,000 full-length human and mouse cDNA sequences"; Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).

CC

CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

CC

DR ENBL; AK001600; BAA91781.1; -

DR ENBL; AL834434; CAD39094.1; ALT_INIT.

DR ENBL; BC001962; AAH01962.1; -

DR Genew; HGNC:20217; C14orf114.

DR InterPro; IPR002562; 3_5_exonuclease.

DR Pfam; PF01612; 3_5_exonuclease; 1.

FT CONFLICT 162 -162 V -> G (IN REF. 2).

FT SEQUENCE 496 AA; 56345 MW; D07693402F85BFB6 CRC64;

QY

Db 366 HLKQOLA 372

347 HLKQOLA 353

Query Match 1.3%; Score 7; DB 1; Length 496;

Best Local Similarity 100.0%; Pred. No. 1.1e+02;

Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
OC Bacteria; Spirochaetes; Spirochaetales; Spirochaetaceae; Treponema.
OX NCBI_TaxID=160;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Nichols;
RX MEDLINE=98332770; PubMed=9665876;
RA Fraser C.M., Norris S.J., Weinstock G.M., White O., Sutton G.G.,
RA Dodson R., Gwin M., Hickey E.K., Clayton R., Ketchum K.A.,
RA Sodergren B., Hardham J.M., McLeod W.P., Salzberg S., Peterson J.,
RA Khalak H., Richardson D., Howell J.K., Chidambaram M., Utterback T.,
RA McDonald L., Artach P., Bowman C., Cotton M.D., Fujii C., Garland S.,
RA Hatch B., Horst K., Roberts K., Sandusky M., Weidman J., Smith H.O.,
RA Venter J.C.;
RT "Complete genome sequence of Treponema pallidum, the syphilis
RT Spirochete";
RL Science 281:375-388(1998).
CC -1- CATALYTIC ACTIVITY: D-glucose 6-phosphate + NADP(+) = D-glucono-
CC 1,5-lactone 6-phosphate + NADPH.
CC -1- PATHWAY: Pentose phosphate + NADPH.
CC -1- SIMILARITY: Belongs to the glucose-6-phosphate dehydrogenase
CC family.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; AE001224; AAC65465.1; -
CC PIR; D71319; D71319.
CC HSP; P11411; 1DPG.
CC TIGR; TP0478; -.
CC InterPro; IPR001282; G6PD.
CC Pfam; PF00479; G6PD; 1.
CC PRINTS; PR00079; G6PDHGNASE.
CC PRODOM; PD001129; G6PD; 1.
CC TIGRFAMs; TIGR00871; zwf; 1.
CC DR PROSITE; PS00069; G6P DEHYDROGENASE; 1.
CC KX Oxidoreductase; NADP; Glucose metabolism; Complete proteome.
CC FT ACT SITE 194 194 BY SIMILARITY.
CC SQ SEQUENCE 515 AA; 58032 MW; 22936A1609EBD798 CRC64;

Query Match 1.3%; Score 7; DB 1; Length 515;
Best Local Similarity 100.0%; Pred. No. 1.2e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 48 QSGALRD 54
Db 241 QSGALRD 247

RESULT 71
TRAS HUMAN STANDARD; PRT; 557 AA.
ID ID TRAS HUMAN STANDARD; PRT; 557 AA.
AC 00043;
DT 10-OCT-2003 (Rel. 42, Created)
DT 10-OCT-2003 (Rel. 42, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE TNF receptor associated factor 5.
GN TRAF5.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A., AND INTERACTION WITH TNFRSF5; TNFRSF8 AND LTR.
RX MEDLINE=98172745; PubMed=9511754;
RA Mizushima S.-I., Fujita M., Ishida T., Azuma S., Kato K., Hirai M.,
RA Otsuka M., Yamamoto T., Inoue J.-I.;
RT "Cloning and characterization of a cDNA encoding the human homolog of
```

```
RESULT 69
CNB4 MOUSE
ID ID CNB4 MOUSE STANDARD; PRT; 496 AA.
AC Q8VEG4;
DT 15-MAR-2004 (Rel. 43, Created)
DT 15-MAR-2004 (Rel. 43, Last sequence update)
DT 15-MAR-2004 (Rel. 43, Last annotation update)
DE Protein C14orf114 homolog.
GN C14ORF114.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Breast tumor;
RX MEDLINE=22388257; PubMed=12477932;
RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
RA Altshul S.F., Zeeberg B., Bustow K.H., Schaefer C.F., Bhat N.K.,
RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hong L.,
RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
RA Brownstein M.J., Udén T.B., Toshiyuki S., Carninci P., Prange C.,
RA Raba S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullaby S.J.,
RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gnanaratne P.H.,
RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Rulyk S.W.,
RA Villalón D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA Fahey J., Helton E., Kettner M., Madan A.C., Rodrigues S., Sanchez A.,
RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA Rodriguez A.C., Grinwood J., Schmutz J., Myers R.M.,
RA Butterfield V.S.N., Krzywinski M.I., Skalska U., Smalls D.E.,
RA Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.;
RT "Generation and initial analysis of more than 15,000 full-length human
RT and mouse cDNA sequences";
RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903 (2002).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; BC018508; AAH18508.1; -
CC MGD; MGI:2145119; C85658.
CC DR InterPro; IPR002562; 3_5_exonuclease.
CC DR Pfam; PF01612; 3_5_exonuclease; 1.
CC SQ SEQUENCE 496 AA; 56785 MW; 8349C98C9800223 CRC64;

Query Match 1.3%; Score 7; DB 1; Length 496;
Best Local Similarity 100.0%; Pred. No. 1.1e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 366 HLKQQLA 372
Db 347 HLKQQLA 353

RESULT 70
ID ID G6PD TREPA STANDARD; PRT; 515 AA.
AC O83491;
DT 15-DEC-1998 (Rel. 37, Created)
DT 15-DEC-1998 (Rel. 37, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Glucose-6-phosphate 1-dehydrogenase (EC 1.1.1.49) (G6PD).
GN ZWF OR TP0478.
OS Treponema pallidum.
```

- RP tumor necrosis factor receptor-associated factor 5 (TRAF5).";
RL Gene 207:135-140(1998).
RN [2]
RP SEQUENCE FROM N.A.
RC TISSUE=Testis;
RX MEDLINE=22388257; PubMed=12477932;
RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA Hopkins R.F., Jordan K., Moore T., Max S.I., Wang J., Hsieh F.,
RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
RA Brownstein M.J., Uddin T.B., Toshiyuki S., Carninci P., Prange C.,
RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullaly S.J.,
RA Bosak S.A., McKean P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
RA Villalón D.K., Muny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA Fahey J., Helton E., Kettman M., Madan A., Rodrigues S., Sanchez A.,
RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,
RA Butterfield Y.S.N., Krzywicki M.I., Skalska U., Smalhus D.E.,
RA Schnerch A., Schein J.E., Jones S.O.M., Marra M.A.;
RT "Generation and initial analysis of more than 15,000 full-length human
RT and mouse cDNA sequences.";
RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
RN [3]
RP SEQUENCE OF 20-557 FROM N.A.
RX MEDLINE=97321041; PubMed=9177772;
RA Nakano H., Shindo M., Yamada K., Yoshida M.C., Santee S.M., Ware C.F.,
RA Jenkins N.A., Gilbert D.J., Yagita H., Copeland N.G., Okumura K.;
RT "Human TNF receptor-associated factor 5 (TRAF5): cDNA cloning,
RT expression and assignment of the TRAF5 gene to chromosome 1q32.";
RL Genomics 42:26-32(1997).
RN [4]
RP REVIEW
RX MEDLINE=21278352; PubMed=11384937;
RA Wajant H., Henkler F., Scheurich P.;
RT "The TNF-receptor-associated factor family: scaffold molecules for
RT cytokine receptors, kinases and their regulators.";
RL Cell. Signal. 13:389-400(2001).
RN [5]
RP REVIEW
RX MEDLINE=21519158; PubMed=11607847;
RA Bradley J.R., Pober J.S.;
RT "Tumor necrosis factor receptor-associated factors (TRAFs).";
RL Oncogene 20:6482-6491(2001).
RN [6]
RP INTERACTION WITH TNFRSF14
RX MEDLINE=97306297; PubMed=9162022;
RA Marsters S.A., Ayres T.M., Skubatch M., Gray C.L., Rothe M.,
RA Ashkenazi A.;
RT "Herpesvirus entry mediator, a member of the tumor necrosis factor
RT receptor (TNFR) family, interacts with members of the TNFR-associated
RT factor family and activates the transcription factors NF-kappaB and
RT AP-1.";
RL J. Biol. Chem. 272:14029-14032(1997).
RN [7]
RP INTERACTION WITH TNFRSF14
RX MEDLINE=97298041; PubMed=91531189;
RA Hsu H., Solovyeff I., Colombero A., Elliott R., Kelley M., Boyle W.J.;
RT "ATAR, a novel tumor necrosis factor receptor family member, signals
RT through TRAF2 and TRAF5.";
RL J. Biol. Chem. 272:13471-13474(1997).
RN [8]
RP INTERACTION WITH MAP3K14
RX MEDLINE=97420762; PubMed=9275204;
RA Song H.Y., Regnier C.H., Kirschning C.J., Goeddel D.V., Rothe M.;
RT "Tumor necrosis factor (TNF)-mediated kinase cascades: bifurcation of
RT nuclear factor-kappaB and c-jun N-terminal kinase (JNK/SAPK) pathways
RT at TNF receptor-associated factor 2.";
RL Proc. Natl. Acad. Sci. U.S.A. 94:9792-9796(1997).
RN [9]
RP INTERACTION WITH TNFRSF8
RX MEDLINE=97152965; PubMed=8998988;
RA Aizawa S., Nakano H., Ishida T., Horie R., Nagai M., Ito K.,
RA Yagita H., Okumura K., Inoue J.-I., Watanabe T.;
RT "Tumor necrosis factor receptor-associated factor (TRAF) 5 and TRAF2
RT are involved in CD30-mediated NF-kappaB activation.";
RL J. Biol. Chem. 272:2042-2045(1997).
RN [10]
RP INTERACTION WITH MAP3K5
RX MEDLINE=98448103; PubMed=9774977;
RA Nishitoh K., Saitoh M., Mochida Y., Takeda K., Nakano H., Rothe M.,
RA Miyazono K., Ichijo H.;
RT "ASK1 is essential for JNK/SAPK activation by TRAF2.";
RL Mol. Cell 2:389-395(1998).
RN [11]
RP INTERACTION WITH TNFRSF5 AND TRAF3
RX MEDLINE=98384149; PubMed=9718306;
RA Pullen S.S., Miller H.G., Everdeen D.S., Dang T.T., Crute J.J.,
RA Kehry M.R.;
RT "CD40-tumor necrosis factor receptor-associated factor (TRAF)
RT interactions: regulation of CD40 signaling through multiple TRAF
RT binding sites and TRAF hetero-oligomerization.";
RL Biochemistry 37:11836-11845(1998).
RN [12]
RP INTERACTION WITH TNFRSF4
RX MEDLINE=98157982; PubMed=9488716;
RA Kawamata S., Hori T., Imura A., Takaori-Kondo A., Uchiyama T.;
RT "Activation of OX40 signal transduction pathways leads to tumor
RT necrosis factor receptor-associated factor (TRAF) 2- and
RT TRAF5-mediated NF-kappaB activation.";
RL J. Biol. Chem. 273:5808-5814(1998).
RN [13]
RP INTERACTION WITH RIPK2
RX MEDLINE=98307936; PubMed=9642260;
RA McCarthy J.V., Ni J., Dixit V.M.;
RT "RIP2 is a novel NF-kappaB-activating and cell death-inducing
RT kinase.";
RL J. Biol. Chem. 273:16968-16975(1998).
RN [14]
RP INTERACTION WITH TNFRSF19
RX MEDLINE=20270246; PubMed=10809768;
RA Eby M.T., Jasmin A., Kumar A., Sharma K., Chaudhary P.M.;
RT "TAJ, a novel member of the tumor necrosis factor receptor family,
RT activates the c-Jun N-terminal kinase pathway and mediates
RT caspase-independent cell death.";
RL J. Biol. Chem. 275:15336-15342(2000).
RN [15]
RP INTERACTION WITH TNFRSF11A
RX MEDLINE=98447691; PubMed=9774460;
RA Kong B.R., Josien R., Lee S.Y., Vologodskaya M., Steinman R.M.,
RA Choi Y.;
RT "The TRAF family of signal transducers mediates NF-kappaB activation
RT by the TRANCE receptor.";
RL J. Biol. Chem. 273:28355-28359(1998).
RN [16]
RP INTERACTION WITH TTRAP
RX MEDLINE=20309820; PubMed=10764746;
RA Eysse S., Declercq W., Ibrahim A., Michiels C., Van Rietachoten J.G.,
RA Dewulf N., de Boer M., Vandenberghe P., Huybrechts D., Remacle J.E.;
RT "TTRAP, a novel protein that associates with CD40, tumor necrosis
RT factor (TNF) receptor-75 and TNF receptor-associated factors (TRAFs),
RT and that inhibits nuclear factor-kappa B activation.";
RL J. Biol. Chem. 275:18586-18593(2000).
RN [17]
RP INTERACTION WITH TNFRSF13B
RX MEDLINE=20341628; PubMed=10880535;
RA Xia X.-Z., Treanor J., Senaldi G., Khare S.D., Boone T., Kelley M.,
RA Theill L.E., Colombero A., Solovyeff I., Lee F., McCabe S., Elliott R.,
RA Miner K., Hawkins N., Guo J., Stolina M., Yu G., Wang J., Delaney J.,
RA Meng S.Y., Boyle W.J., Hsu H.;
RT "TACI is a TRAP-interacting receptor for TAIL-1, a tumor necrosis
RT factor family member involved in B cell regulation.";
RL J. Exp. Med. 192:137-143(2000).


```

RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=CT18;
RX MEDLINE=21534947; PubMed=11677608;
RA Parkhill J., Dougan G., James K.D., Thomson N.R., Pickard D., Wain J.,
RA Churcher C., Mungall K.L., Bentley S.D., Holden M.T.G., Sebahia M.,
RA Baker S., Basham D., Brooks K., Chillingworth T., Connor P.,
RA Cronin A., Davis P., Davies R.M., Dowd L., White N., Farrar J.,
RA Feltwell T., Hamlin N., Haque A., Hien T.T., Holroyd S., Jagels K.,
RA Krogh A., Larsen T.S., Leather S., Moule S., O'Gaora P., Parry C.,
RA Quail M.A., Rutherford K., Simmonds M., Skelton J., Stevens K.,
RA Whithead S., Barrall B.G.;
RT "Complete genome sequence of a multiple drug resistant Salmonella
RT enterica serovar Typhi CT18."
RL Nature 413:848-852(2001).
RN [4]
RP SEQUENCE FROM N.A.
RC STRAIN=TY2 / ATCC 700931;
RX MEDLINE=22531367; PubMed=12644504;
RA Deng W., Liou S.-R., Plunkett G. III, Mayhew G.F., Rose D.J.,
RA Burland V., Kodyanni V., Schwartz D.C., Blattner F.R.;
RT "Comparative genomics of Salmonella enterica serovar Typhi strains Ty2
RT and CT18."
RL J. Bacteriol. 185:2330-2337(2003).
RN -1- PATHWAY: Vi polysaccharide biosynthesis.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; X67785; CAA47994.1; -
DR EMBL; D14156; BAA03195.1; -
DR EMBL; AL627283; C8D06776.1; -
DR EMBL; AB016848; AB071802.1; -
DR PIR; E36892; E36892.
DR PIR; S28492; S28492.
DR InterPro; IPR001296; Glyco_transf_1.
DR Pfam; PF00534; Glycos_transf_1; 1.
KW Complete proteome.
FT CONFLICT 290 290 S -> G (IN REF. 1 AND 4).
FT CONFLICT 362 363 KH -> ND (IN REF. 1).
FT CONFLICT 363 363
SQ SEQUENCE 578 AA; 65009 MW; 8D420563D868C189 CRC64;

Query Match 1.3%; Score 7; DB 1; Length 578;
Best Local Similarity 100.0%; Pred. No. 1.3e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 170 EKKLAAL 176
DB 5 EKKLAAL 11

RESULT 74
Y085 CHLTR STANDARD; PRT; 579 AA.
AC G84087;
DT 30-MAY-2000 (Rel. 39, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Hypothetical protein CT085.
GN CT085.
OS Chlamydia trachomatis.
OC Bacteria; Chlamydiae; Chlamydiales; Chlamydiaceae; Chlamydia.
OX NCBI_TaxID=813;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=D/UN-3/CX;
RX MEDLINE=9900809; PubMed=97841136;
RA Stephens R.S., Kalman S., Lammel C.J., Fan J., Marathe R., Aravind L.,

```

```

RA Mitchell W.P., Olinger L., Tatusov R.L., Zhao Q., Koonin E.V.,
RA Davis R.W.;
RT "Genome sequence of an obligate intracellular pathogen of humans:
RT Chlamydia trachomatis."
RL Science 282:754-759(1998).
CC -1- SIMILARITY: Belongs to the ubiD family.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AB001283; AAC67676.1; -
DR PIR; E71557; E71557.
DR InterPro; IPR002830; carboxylase.
DR Pfam; PF01977; UbiD; 1.
DR TIGRfams; TIGR00148; TIGR00148; 1.
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 579 AA; 65354 MW; E2512FC314216315 CRC64;

Query Match 1.3%; Score 7; DB 1; Length 579;
Best Local Similarity 100.0%; Pred. No. 1.3e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 486 ERRPFGP 492
DB 268 ERRPFGP 274

RESULT 75
Y454 CHLCV STANDARD; PRT; 581 AA.
AC Q34023;
DT 30-MAY-2000 (Rel. 39, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 10-OCT-2001 (Rel. 42, Last annotation update)
DE Hypothetical protein CCA00454 (ORF).
GN CCA00454.
OS Chlamydia caviae.
OC Bacteria; Chlamydiae; Chlamydiales; Chlamydiaceae; Chlamydia.
OX NCBI_TaxID=83557;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=GPIC;
RX MEDLINE=97426043; PubMed=9282747;
RA Hsia R.C., Pannekoek Y., Ingerowski E., Bavoil P.M.;
RT "Type III secretion genes identify a putative virulence locus of
RT Chlamydia."
RL Mol. Microbiol. 25:351-359(1997).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=GPIC;
RX MEDLINE=22569155; PubMed=12682364;
RA Read T.D., Myers G.S.A., Brunham R.C., Nelson W.C., Paulsen I.T.,
RA Heidelberg J., Holtapple E., Khouri H., Federova N.B., Carty H.A.,
RA Umayam L.A., Haft D.H., Peterson J., Beanan M.J., White O.,
RA Salzberg S.L., Hsia R.-C., McClarty G., Rank R.G., Bavoil P.M.,
RA Fraser C.M.;
RT "Genome sequence of Chlamydia caviae (Chlamydia psittaci GPIC):
RT examining the role of niche-specific genes in the evolution of the
RT Chlamydiaceae."
RL Nucleic Acids Res. 31:2134-2147(2003).
CC -1- SIMILARITY: Belongs to the ubiD family.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/

```

```

CC or send an email to license@isb-sib.ch].
CC -----
DR EMBL; U88070; AAB71513.1; -.
DR EMBL; AE016995; AAP05200.1; -.
DR TIGR; CCA000454; -.
DR InterPro; IPR002830; carboxylase.
DR Pfam; PF01977; UbiD; 1.
DR TIGRFAMs; TIGR00148; TIGR00148; 1.
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 581 AA; 66207 MW; 8D51F340DD99A2DC CRC64;

Query Match
Best Local Similarity 100.0%; Score 7; DB 1; Length 581;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 486 ERRPEGP 492
DB 268 ERRPEGP 274

RESULT 76
DI70 HUMAN
ID DI70 HUMAN STANDARD; PRT; 585 AA.
AC Q16204;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 15-MAR-2004 (Rel. 43, Last annotation update)
DE H4 protein.
GN D10S170.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Thyroid;
RX MEDLINE=94336206; PubMed=8058316;
RA Grieco M., Carriato A., Santoro M., Fusco A., Melillo R.M., Vecchio G.;
RT "Cloning and characterization of H4 (D10S170), a gene involved in RET
RT rearrangements in vivo.";
RL Oncogene 9:2531-2535(1994).
CC -!- SURCELLULAR LOCATION: May be a cytoskeletal protein.
CC -!- TISSUE SPECIFICITY: Ubiquitously expressed.
CC -!- DOMAIN: the protein has mostly an alpha helical conformation
CC similar to myosin heavy-chain tail that might adopt a coiled-coil
CC conformation.
CC -!- DISEASE: ABOUT 20% OF HUMAN THYROID PAPILLARY CARCINOMAS (PACT)
CC SEEM TO BE CAUSED BY A REARRANGEMENT INV(10)(Q11.2Q21) INVOLVING
CC RET AND H4(D10S170) THAT GENERATES THE RET/PTC1 ONCOGENE.
CC -!- SIMILARITY: TO C.ELEGANS T09B9.4.

This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).

EMBL; S72869; AAC60637.1; -.
DR PIR; I58403; I58403.
DR MIM; 601985; -.
DR MIM; 188550; -.
DR GO; GO:0005200; F:structural constituent of cytoskeleton; TAS.
KW Proto-oncogene; Chromosomal translocation; SH3-binding; Repeat.
FT DOMAIN 17 20 POLY-SER.
FT DOMAIN 32 44 POLY-GLY.
FT DOMAIN 440 448 POLY-PRO.
FT DOMAIN 106 235 5 X 29 AA TANDEM REPEATS.
FT REPEAT 106 134 1.
FT REPEAT 135 163 2.
FT REPEAT 164 192 3.
FT REPEAT 193 206 4 (APPROXIMATE).

```

```

FT REPEAT 207 235 5.
FT SITE 442 451 SH3-BINDING (POTENTIAL).
SQ SEQUENCE 585 AA; 65916 MW; B3FD9C81DA72AA8A CRC64;

Query Match
Best Local Similarity 100.0%; Score 7; DB 1; Length 585;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 404 KIKKLE 410
DB 176 KIKKLE 182

RESULT 77
YM72 YEAST
ID YM72 YEAST STANDARD; PRT; 590 AA.
AC Q05021;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Hypothetical 67.6 kDa protein in MRPL44-MTF1 intergenic region.
GN YMR227C OR YN9599.09C.
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
OX NCBI_TaxID=4932;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=S288C / AB972;
RX MEDLINE=97313268; PubMed=9169872;
RA Bowman S., Churcher C.M., Badcock K., Brown D., Chillingworth T.,
RA Connor R., Dedman K., Devlin K., Gentles S., Hamlin N., Hunt S.,
RA Jagels K., Lye G., Moule S., Odell C., Pearson D., Rajandream M.A.,
RA Rice P., Skelton J., Walsh S., Whitehead S., Barrell B.G.;
RT "The nucleotide sequence of Saccharomyces cerevisiae chromosome
RT XIII.";
RL Nature 387:90-93(1997).

This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).

EMBL; Z49933; CAA90198.1; -.
DR PIR; S57594; S57594.
DR Germonline; 142902; -.
DR TRANSFAC; T03088; -.
DR SGD; S0004840; TAF67.
DR GO; GO:0005634; C:nucleus; IDA.
DR GO; GO:0000114; P:G1-specific transcription in mitotic cell c.; IPI.
DR InterPro; IPR006751; TAFI55_N.
DR Pfam; PF04658; TAFI55_N; 1.
KW Hypothetical protein; Coiled coil.
FT DOMAIN 50 53 POLY-LYS.
FT DOMAIN 200 203 POLY-GLU.
FT DOMAIN 368 373 POLY-GLU.
FT DOMAIN 413 421 POLY-ASP.
FT DOMAIN 517 531 POLY-GLU.
FT DOMAIN 427 549 COILED COIL (POTENTIAL).
SQ SEQUENCE 590 AA; 67555 MW; C014E7419B0B1C61 CRC64;

Query Match
Best Local Similarity 100.0%; Score 7; DB 1; Length 590;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 406 IKKLEKE 412
DB 479 IKKLEKE 485

```

```

RESULT 78
DNJM_MYCGE
ID DNJM_MYCGE STANDARD; PRT; 601 AA.
AC P4742; Q49288;
DT 01-FEB-1996 (Rel. 33, Created)
DE 01-FEB-1996 (Rel. 33, Last sequence update)
DE 16-OCT-2001 (Rel. 40, Last annotation update)
DE DnaJ-like protein MG200.
GN MG200.
OS Mycoplasma genitalium.
OC Bacteria; Firmicutes; Mollicutes; Mycoplasmataceae; Mycoplasma.
OX NCBI_TaxID=2097;
RN [1]
SEQUENCE FROM N.A.
RC STRAIN=ATCC 33530 / G-37;
RX MEDLINE=96026346; PubMed=7569993;
RA Fraser C.M., Gocayne J.D., White O., Adams M.D., Clayton R.A.,
RA Fleischmann J.L., Weidman J.F., Small K.V., Sandusky M., Fuhrmann J.L.,
RA Nguyen D.T., Ufferback T.R., Saudek D.M., Phillips C.A., Merrick J.M.,
RA Tomb J.-F., Dougherty B.A., Boff K.P., Hu P.-C., Lucier T.S.,
RA Peterson S.N., Smith H.O., Hutchison C.A. III, Venter J.C.;
RT "The minimal gene complement of Mycoplasma genitalium."
RL Science 270:397-403(1995).
RN [2]
SEQUENCE OF 281-409 FROM N.A.
RC STRAIN=ATCC 33530 / G-37;
RX MEDLINE=94075230; PubMed=8253680;
RA Peterson S.N., Hu P.-C., Boff K.P., Hutchison C.A. III;
RT "A survey of the Mycoplasma genitalium genome by using random
RT sequencing."
RL J. Bacteriol. 175:7918-7930(1993).
CC -!- SIMILARITY: Contains 1 J domain.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; U39699; AAC71418.1; -
DR EMBL; U02163; AAD12445.1; -
DR PIR; A64222; A64222.
DR HSSP; P08622; LBQZ.
DR TIGR; MG200; -.
DR InterPro; IPR001623; DnaJ_N.
DR InterPro; IPR003095; Hsp_DnaJ.
DR Pfam; PF00226; DnaJ; 1.
DR PRINTS; PR00625; DnaJPROTEIN.
DR SMART; SM00271; DnaJ; 1.
DR PROSITE; PS00636; DnaJ_1; 1.
DR PROSITE; PS00076; DnaJ_2; 1.
DR KW Hypothetical protein; Chaperone; Complete proteome.
FT DOMAIN 5 77 J-DOMAIN.
SQ SEQUENCE 601 AA; 68537 MW; F9FAE352E341D093 CRC64;
Query Match 1.3%; Score 7; DB 1; Length 601;
Best Local Similarity 100.0%; Pred. No. 1.4e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 106 EPEPTPV 112
Db 215 EPEPTPV 221
|||||
RESULT 79
SP2_HUMAN
ID SP2_HUMAN STANDARD; PRT; 606 AA.
AC Q02086;
DT 01-FEB-1995 (Rel. 31, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)

```

10-OCT-2003 (Rel. 42, Last annotation update)

Transcription factor Sp2.

SP2 OR KIA0048.

OS Homo sapiens (Human).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

NCBI_TaxID=9606;

[1]

SEQUENCE FROM N.A.

RC TISSUE=Bone marrow;

RX MEDLINE=96051398; PubMed=7584044;

RA Nomura N., Nagase T., Miyajima N., Suzuki T., Tanaka A., Sato S., Seki N., Kawatabayashi Y., Ishikawa K.-I., Tabata S.;

RT "Prediction of the coding sequences of unidentified human genes. II. The coding sequences of 40 new genes (K1A0041-K1A0080) deduced by analysis of cDNA clones from human cell line KG-1.";

RNA Res. 1:223-229(1994).

[2]

SEQUENCE FROM N.A.

RC TISSUE=Lymph, and Spleen;

RX MEDLINE=22388257; PubMed=12477932;

RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G., Schuler G.D., Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schaefer C.P., Bhat N.K., Altschul S.F., Zeeberg B., Buetow K.H., Moore T., Max S.I., Wang J., Hsieh P., Hopkins R.F., Jordan H., Moore M., Farmer A.A., Rubin G.M., Hong L., Stapleton M., Soares M.B., Bonaldo M.P., Casavant T.L., Scheetz T.E., Brownstein M.J., Ustin T.B., Toshynski S., Carninci P., Prange C., Raba S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullany S.J., Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H., Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W., Villalón D.K., Wuzny D.M., Sodergren E.J., Lu X., Gibbs R.A., Fahey J., Helton E., Kettman M., Madan A., Rodriguez S., Sanchez A., Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G., Rada S.S., Touchman J.W., Green E.D., Dickson M.C., Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M., Butlerfield Y.S.N., Krawinski M.I., Skalska U., Smailus D.E., Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.;

RT "Generation and initial analysis of more than 15,000 full-length human and mouse cDNA sequences.";

RNA Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).

[3]

SEQUENCE OF 112-606 FROM N.A.

RX MEDLINE=93024366; PubMed=1341900;

RA Kingsley C., Winoto A.;

RT "Cloning of GT box-binding proteins: a novel Sp1 multigene family regulating T-cell receptor gene expression.";

RL Mol. Cell. Biol. 12:4251-4261(1992).

CC -!- FUNCTION: BINDS TO GC BOX PROMOTERS ELEMENTS AND SELECTIVELY ACTIVATES MENA SYNTHESIS FROM GENES THAT CONTAIN FUNCTIONAL RECOGNITION SITES.

CC -!- SUBCELLULAR LOCATION: Nuclear.

CC -!- SIMILARITY: BELONGS TO THE SP1 FAMILY OF C2H2-TYPE ZINC-FINGER PROTEINS.

CC -!- SIMILARITY: Contains 3 C2H2-type zinc fingers.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See http://www.isb-sib.ch/announce/ or send an email to license@isb-sib.ch).

CC -----

EMBL; D28588; BAA05923.1; -

EMBL; BC016680; AAH16680.1; -

EMBL; BC033814; AAH33814.1; -

EMBL; M97190; AAA36629.1; -

PIR; A44489; A44489.

HSSP; P08047; 1SP2.

DR TRANSFAC; T02356; -

DR Genew; HGNC:11207; SP2.

MIN; 601801; -

DR GO: GO:0003702; F:RNA polymerase II transcription factor acti. . .; TAS.
DR GO: GO:0006955; P:immune response; TAS.
DR GO: GO:0006357; P:regulation of transcription from Pol II pro. . .; TAS.
DR InterPro: IPR007087; Znf_C2H2.
DR Pfam: PF00096; ZF-C2H2; 3.
DR ProDom: PD000003; Znf_C2H2; 2.
DR SMART: SMO0355; Znf_C2H2; 3.
DR PROSITE; PS00028; ZINC_FINGER_C2H2_1; 3.
DR PROSITE; PS00157; ZINC_FINGER_C2H2_2; 3.
KW Transcription regulation; Activator; Zinc-finger; Metal-binding;
KW DNA-binding; Nuclear protein; Repeat.
FT ZN_FING 518 542 C2H2-TYPE 1.
FT ZN_FING 548 572 C2H2-TYPE 2.
FT ZN_FING 578 600 C2H2-TYPE 3.
SQ SEQUENCE 606 AA; 64153 MW; A27CGD460D36E186 CRC64;

Query Match 1.3%; Score 7; DB 1; Length 606;
Best Local Similarity 100.0%; Pred. No. 1.4e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 268 IQLQMEQ 274
DB 475 IQLQMEQ 481

RESULT 80
ID DNAK_BUCBP STANDARD; PRT; 638 AA.
AC Q8K9Y8;
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Chaperone protein dnaK (Heat shock protein 70) (Heat shock 70 kDa protein) (HSP70).
GN DNAK OR BUSG146.
OS Buchnera aphidicola (subsp. Schizaphis graminum).
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Buchnera.
OX NCBI_TaxID=98794;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=22084549; PubMed=12089438;
RA Tamas I., Klasson L., Canbaeck B., Naeslund A.K., Eriksson A.-S.,
RA Wernegreen J.J., Sandstrom J.P., Moran N.A., Andersson S.G.E.;
RT "50 million years of genomic stasis in endosymbiotic bacteria.";
RL Science 296:2376-2379(2002).
CC -1- FUNCTION: Acts as a chaperone (By similarity).
CC -1- INDUCTION: By stress conditions e.g. heat shock (By similarity).
CC -1- SIMILARITY: Belongs to the heat shock protein 70 family.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; AE014090; AAM67714.1; .
CC HAMAP; MF 00332; .
CC InterPro: IPR001023; Hsp70.
CC Pfam; PF00012; HSP70; 1.
CC PRINTS; PR00301; HEATSHOCK70.
CC ProDom; PD000089; HSP70; 1.
CC PROSITE; PS00297; HSP70_1; 1.
CC PROSITE; PS00329; HSP70_2; 1.
CC PROSITE; PS01036; HSP70_3; 1.
KW Chaperone; ATP-binding; Heat shock; Phosphorylation;
KW Complete proteome. 199
FT MOD_RES 199 PHOSPHORYLATION (AUTO-) (BY SIMILARITY).
SQ SEQUENCE 638 AA; 70344 MW; 541CCFE2581F043F CRC64;

Query Match 1.3%; Score 7; DB 1; Length 638;
Best Local Similarity 100.0%; Pred. No. 1.4e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY

Best Local Similarity 100.0%; Pred. No. 1.4e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 222 RSKLES 228
DB 302 RSKLES 308

RESULT 81
ID DNAK_BUCBP STANDARD; PRT; 638 AA.
AC PS9565;
DT 10-OCT-2003 (Rel. 42, Created)
DT 10-OCT-2003 (Rel. 42, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Chaperone protein dnaK (Heat shock protein 70) (Heat shock 70 kDa protein) (HSP70).
GN DNAK OR BBP142.
OS Buchnera aphidicola (subsp. Baizongia pistaciae).
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Buchnera.
OX NCBI_TaxID=135842;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=22426901; PubMed=12522265;
RA Van Ham R.C.H.J., Kamerbeek J., Palacios C., Rausell C., Abascal F.,
RA Bastolla U., Fernandez J.M., Jimenez L., Postigo M., Silva F.J.,
RA Tamames J., Viguera E., Latorre A., Valencia A., Moran F., Moya A.;
RT "Reductive genome evolution in Buchnera aphidicola.";
RL Proc. Natl. Acad. Sci. U.S.A. 100:581-586(2003).
CC -1- FUNCTION: Acts as a chaperone (By similarity).
CC -1- INDUCTION: By stress conditions e.g. heat shock (By similarity).
CC -1- SIMILARITY: Belongs to the heat shock protein 70 family.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; AE014016; AAO26876.1; .
CC HAMAP; MF 00332; .
CC InterPro: IPR001023; Hsp70.
CC Pfam; PF00012; HSP70; 1.
CC PRINTS; PR00301; HEATSHOCK70.
CC ProDom; PD000089; HSP70; 1.
CC PROSITE; PS00297; HSP70_1; 1.
CC PROSITE; PS00329; HSP70_2; 1.
CC PROSITE; PS01036; HSP70_3; 1.
KW Chaperone; ATP-binding; Heat shock; Phosphorylation;
KW Complete proteome. 199
FT MOD_RES 199 PHOSPHORYLATION (AUTO-) (BY SIMILARITY).
SQ SEQUENCE 638 AA; 70368 MW; 9C4F87571EE0AD82 CRC64;

Query Match 1.3%; Score 7; DB 1; Length 638;
Best Local Similarity 100.0%; Pred. No. 1.4e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 222 RSKLES 228
DB 302 RSKLES 308

RESULT 82
ID DNAK_PSESG STANDARD; PRT; 638 AA.
AC Q9WMG9;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Chaperone protein dnaK (Heat shock protein 70) (Heat shock 70 kDa

```

DE protein) (HSP70).
GN DNAX.
OS Pseudomonas syringae (pv. glycinea).
OC Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales;
OC Pseudomonadaceae; Pseudomonas.
OX NCBI_TaxID=318;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=PG4180;
RA MEDLINE=99407915; PubMed=10478477;
RX Keith L.M.W., Partridge J.E., Bender C.L.;
RT "dnak and the heat stress response of Pseudomonas syringae pv.
RL glycinea.";
RT Mol. Plant Microbe Interact. 12:563-574(1999).
CC -!- FUNCTION: Acts as a chaperone (By similarity).
CC -!- INDUCTION: By stress conditions e.g. heat shock (By similarity).
CC -!- SIMILARITY: Belongs to the heat shock protein 70 family.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; AF135163; AAD31868.1; -.
DR HSP; P04475; IDK1.
DR HAMAP; MF_00332; -.
DR InterPro; IPR001023; Hsp70.
DR Pfam; PF00012; HSP70; 1.
DR PRINTS; PR00301; HEATSHOCK70.
DR ProDom; PD000089; Hsp70; 1.
DR PROSITE; PS00297; HSP70_1; 1.
DR PROSITE; PS00329; HSP70_2; 1.
DR PROSITE; PS01036; HSP70_3; 1.
DR Chaperone; ATP-binding; Heat shock; Phosphorylation.
KW Chaperone; ATP-binding; Heat shock; Phosphorylation.
FT MOD RES 199 199 PHOSPHORYLATION (AUTO-) (BY SIMILARITY).
FT MOD RES 199 199 PHOSPHORYLATION (AUTO-) (BY SIMILARITY).
SQ SEQUENCE 638 AA; 68897 MW; DE2C34D287FC21B CRC64;

Query Match 1.3%; Score 7; DB 1; Length 638;
Best Local Similarity 100.0%; Pred. No. 1.4e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 222 RSKLES 228
DB 302 RSKLES 308
|||||
|||||

RESULT 83
ID DNAX_PSESM STANDARD; PRT; 638 AA.
AC Q87WF0;
DT 10-OCT-2003 (Rel. 42, Created)
DT 10-OCT-2003 (Rel. 42, Last sequence update)
DE Chaperone protein dnak (Heat shock protein 70) (Heat shock 70 kDa
DE protein) (HSP70).
GN DNAX OR PSPT04505.
OS Pseudomonas syringae (pv. tomato).
OC Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales;
OC Pseudomonadaceae; Pseudomonas.
OX NCBI_TaxID=323;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=DC3000;
RX MEDLINE=22834015; PubMed=12928499;
RA Buell C.R., Joardar V., Lindeberg M., Selengut J., Paulsen I.T.,
RA Gwinn M.L., Dodson R.J., Deboy R.T., Durkin A.S., Kolonay J.F.,
RA Madupu R., Daugherty S., Brinkac L., Beanan M.J., Haft D.H.,
RA Nelson W.C., Davidson T., Zafar N., Zhou L., Liu J., Yuan Q.,
RA Khouri H., Fedorova N., Tran B., Russell D., Berry K., Utterback T.,
RA Van Aken S.E., Feldblyum T.V., D'Ascenzo M., Deng W.-L., Ramos A.R.,

```

```

RA Alfano J.R., Cartinhour S., Chatterjee A.K., Delaney T.P.,
RA Lazarowitz S.G., Martin G.B., Schneider D.J., Tang X., Bender C.L.,
RA White O., Fraser C.M., Collmer A.;
RT "The complete genome sequence of the Arabidopsis and tomato pathogen
RT Pseudomonas syringae pv. tomato DC3000."; 10186(2003).
RL Proc. Natl. Acad. Sci. U.S.A. 100:10181-10186(2003).
CC -!- FUNCTION: Acts as a chaperone (By similarity).
CC -!- INDUCTION: By stress conditions e.g. heat shock (By similarity).
CC -!- SIMILARITY: Belongs to the heat shock protein 70 family.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; AE016872; AA057953.1; -.
DR TIGR; PSPT04505; -.
DR HAMAP; MF_00332; -.
DR InterPro; IPR001023; Hsp70.
DR Pfam; PF00012; HSP70; 1.
DR PRINTS; PR00301; HEATSHOCK70.
DR PROSITE; PS00297; HSP70_1; 1.
DR PROSITE; PS00329; HSP70_2; 1.
DR PROSITE; PS01036; HSP70_3; 1.
DR Chaperone; ATP-binding; Heat shock; Phosphorylation;
KW Chaperone; ATP-binding; Heat shock; Phosphorylation;
FT MOD RES 199 199 PHOSPHORYLATION (AUTO-) (BY SIMILARITY).
FT MOD RES 199 199 PHOSPHORYLATION (AUTO-) (BY SIMILARITY).
SQ SEQUENCE 638 AA; 68755 MW; E6504919AA678BCC CRC64;

Query Match 1.3%; Score 7; DB 1; Length 638;
Best Local Similarity 100.0%; Pred. No. 1.4e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 222 RSKLES 228
DB 302 RSKLES 308
|||||
|||||

RESULT 84
ID UVRB_CLOPE STANDARD; PRT; 659 AA.
AC Q46323;
DT 01-NOV-1997 (Rel. 35, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DE UvrABC system protein B (UvrB protein) (Excinuclease ABC subunit B).
GN UVRB OR CPE0303.
OS Clostridium perfringens.
OC Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;
OC Clostridium.
OX NCBI_TaxID=1502;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=13 / Type A;
RX MEDLINE=21664373; PubMed=11792842;
RA Shiba T., Ohtani K., Hirakawa H., Ohshima K., Yamashita A.,
RA Shiba T., Ogasawara N., Hattori M., Kuhara S., Hayashi H.;
RT "Complete genome sequence of Clostridium perfringens, an anaerobic
RT flesh-eater.";
RL Proc. Natl. Acad. Sci. U.S.A. 99:996-1001(2002).
RN [2]
RP SEQUENCE OF 69-155 FROM N.A.
RC STRAIN=CPN50;
RX MEDLINE=96033407; PubMed=759358;
RA Katayama S.-I., Dupuy B., Garnier T., Cole S.T.;
RT "Rapid expansion of the physical and genetic map of the chromosome of
RT Clostridium perfringens CPN50.";
RJ. Bacteriol. 177:5680-5685(1995).
CC -!- FUNCTION: The UvrABC repair system catalyzes the recognition and
CC processing of DNA lesions. A damage recognition complex composed

```

of 2 uvrA and 2 uvrB subunits scans DNA for abnormalities. Upon binding of the uvrA(2)B(2) complex to a putative damaged site, the DNA wraps around one uvrB monomer. DNA wrap is dependent on ATP binding by uvrB and probably causes local melting of the DNA helix, facilitating insertion of uvrB beta-hairpin between the DNA strands. Then uvrB probes one DNA strand for the presence of a lesion. If a lesion is found the uvrA subunits dissociate and the uvrB-DNA preincision complex is formed. This complex is subsequently bound by uvrC and the second uvrB is released. If no lesion is found, the DNA wraps around the other uvrB subunit that will check the other strand for damage (By similarity).

-!- SUBUNIT: Forms a heterotrimer with uvrA during the search for lesions. Interacts with uvrC in an incision complex (By similarity).

-!- SUBCELLULAR LOCATION: Cytoplasmic (By similarity).

-!- DOMAIN: The beta-hairpin motif is involved in DNA binding (By similarity).

-!- SIMILARITY: Belongs to the uvrB family.

-!- SIMILARITY: Contains 1 UVR domain.

This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

EMBL; AP003186; BAB80009.1; --
EMBL; X86531; CAAG0243.1; --
HSSP; P56981; ID9X.
HAMAP; MF_00204; --; 1.
InterPro; IPR001410; DEAD.
InterPro; IPR001650; Helicase_C.
InterPro; IPR004807; UvrB.
InterPro; IPR001943; UvrB/C.
Pfam; PF00271; Helicase_C; 1.
Pfam; PF02151; UVR; 1.
SMART; SM00487; DEXDC; 1.
SMART; SM00490; HELIC_C; 1.
TIGRFAMs; TIGR00631; uvrB; 1.
PROSITE; PS50151; UVR; 1.
SOS response; Excision nuclease; DNA repair; DNA recombination; DNA excision; ATP-binding; Complete proteome.
NP_BIND 38 45
DOMAIN 91 114 BETA-HAIRPIN.
FT DOMAIN 622 657 UVR.
FT CONFLICT 69 75 AAQLCAE -> GSALIVK (IN REF. 2).
SQ SEQUENCE 659 AA; 75523 MW; 4F0D94B3C647B01A CRC64;

Query Match 1.3%; Score 7; DB 1; Length 659;
Best Local Similarity 100.0%; Pred.No. 1.5e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 295 KLIQVE 301
|||||
DB 623 KLIQVE 629

RESULT 85
UVRB METAC
ID UVRB METAC STANDARD; PRT; 670 AA.
AC Q8TK53;
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE UVRABC system protein B (UvrB protein) [Excision nuclease ABC subunit B].
GN UVRB OR MA3323.
OS Methanosarcina acetivorans.
OC Archaea; Euryarchaeota; Methanomicrobia; Methanosarcinales;
OC Methanosarcinaceae; Methanosarcina.
OX NCBI_taxid=2214;
RN [1]

SEQUENCE FROM N.A.
STRAIN-C2A / ATCC 35395 / DSM 2834;
MEDLINE-21929760; PubMed-1193238;
Galagan J.E., Nushbaum C., Roy A., Endrizzi M.G., MacDonald P., FitzRugh W., Culshaw S., Engels R., Smirnov S., Atnoor D., Brown A., Allen N., Naylor J., Stange-Thomann N., DeAtrellano K., Johnson R., Linton L., McEwan P., McKernan K., Talamas J.E., Tirrell A., Ye W., Zimmer A., Barber R.D., Cann I., Graham D.E., Grahame D.A., Guss A.M., Hedderich R., Ingram-Smith C., Kuettnner H.C., Krzycki J.A., Leigh J.A., Li W., Liu J., Mukhopadhyay B., Reeve J.N., Smith K., Springer T.A., Umayam L.A., White O., White R.H., de Macario E.C., Parry J.G., Jarrell K.F., Jing H., Macario A.J.L., Paulsen I., Pritchett M., Sowers K.R., Swanson R.V., Zinder S.H., Lander E., Metcalf M.W., Birren B.;
RT "The genome of Methanosarcina acetivorans reveals extensive metabolic and physiological diversity.";
RL Genome Res. 12:532-542(2002).

-!- FUNCTION: The UVRABC repair system catalyzes the recognition and processing of DNA lesions. A damage recognition complex composed of 2 uvrA and 2 uvrB subunits scans DNA for abnormalities. Upon binding of the uvrA(2)B(2) complex to a putative damaged site, the DNA wraps around one uvrB monomer. DNA wrap is dependent on ATP binding by uvrB and probably causes local melting of the DNA helix, facilitating insertion of uvrB beta-hairpin between the DNA strands. Then uvrB probes one DNA strand for the presence of a lesion. If a lesion is found the uvrA subunits dissociate and the uvrB-DNA preincision complex is formed. This complex is subsequently bound by uvrC and the second uvrB is released. If no lesion is found, the DNA wraps around the other uvrB subunit that will check the other strand for damage (By similarity).

-!- SUBUNIT: Forms a heterotrimer with uvrA during the search for lesions. Interacts with uvrC in an incision complex (By similarity).

-!- SUBCELLULAR LOCATION: Cytoplasmic (By similarity).

-!- DOMAIN: The beta-hairpin motif is involved in DNA binding (By similarity).

-!- SIMILARITY: Belongs to the uvrB family.

-!- SIMILARITY: Contains 1 UVR domain.

This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

EMBL; AE011036; AAM06692.1; --
HAMAP; MF_00204; --; 1.
InterPro; IPR001410; DEAD.
InterPro; IPR001650; Helicase_C.
InterPro; IPR004807; UvrB.
InterPro; IPR001943; UvrB/C.
Pfam; PF00271; Helicase_C; 1.
Pfam; PF02151; UVR; 1.
SMART; SM00487; DEXDC; 1.
SMART; SM00490; HELIC_C; 1.
TIGRFAMs; TIGR00631; uvrB; 1.
PROSITE; PS50151; UVR; 1.
SOS response; Excision nuclease; DNA repair; DNA recombination; DNA excision; ATP-binding; Complete proteome.
NP_BIND 64 71
DOMAIN 117 140 BETA-HAIRPIN.
FT DOMAIN 631 665 UVR.
SQ SEQUENCE 670 AA; 77788 MW; 4BB8ED22CB97188B CRC64;

Query Match 1.3%; Score 7; DB 1; Length 670;
Best Local Similarity 100.0%; Pred.No. 1.5e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 406 IKKLEKE 412
|||||
DB 660 IKKLEKE 666

```

RESULT 86
ID UVRB METWA STANDARD; PRT; 670 AA.
AC QSPZ9;
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DE UvrABC system protein B (UvrB protein) (Excinuclease ABC subunit B).
GN UVRB OR MM3289.
OS Methanosarcina mazei (Methanosarcina frisia).
OC Archaea; Euryarchaeota; Methanomicrobia; Methanosarcinales;
OC Methanosarcinaceae; Methanosarcina.
OX NCBI_TaxID=2209;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Goel / G01 / ATCC BAA-199 / DSM 3647 / OCM 88;
RX MEDLINE=22120827; PubMed=1212524;
RA Deppeiner U., Johann A., Hartesch T., Markl R., Schmitz R.A.,
RA Martinez-Arias R., Henne A., Wietzer A., Baeumer S., Jacobi C.,
RA Bruggemann H., Lienard T., Christmann A., Boemcke M., Steckel S.,
RA Bhattacharyya A., Lykidis A., Overbeek R., Klenk H.-P., Gunsalus R.P.,
RA Fritz H.-J., Gottschalk G.;
RT "The genome of Methanosarcina mazei: evidence for lateral gene
RT transfer between Bacteria and Archaea.";
RL J. Mol. Microbiol. Biotechnol. 4:453-461(2002).
CC -!- FUNCTION: The UvrABC repair system catalyzes the recognition and
CC processing of DNA lesions. A damage recognition complex composed
CC of 2 uvrA and 2 uvrB subunits scans DNA for abnormalities. Upon
CC binding of the uvrA(2)B(2) complex to a putative damaged site, the
CC DNA wraps around one uvrB monomer. DNA wrap is dependent on ATP
CC binding by uvrB and probably causes local melting of the DNA
CC helix, facilitating insertion of uvrB beta-hairpin between the DNA
CC strands. Then uvrB probes one DNA strand for the presence of a
CC lesion. If a lesion is found the uvrA subunits dissociate and the
CC uvrB-DNA preincision complex is formed. This complex is
CC subsequently bound by uvrC and the second uvrB is released. If no
CC lesion is found, the DNA wraps around the other uvrB subunit that
CC will check the other strand for damage (By similarity).
CC -!- SUBUNIT: Forms a heterotetramer with uvrA during the search for
CC lesions. Interacts with uvrC in an incision complex (By
CC similarity).
CC -!- SUBCELLULAR LOCATION: Cytoplasmic (By similarity).
CC -!- DOMAIN: The beta-hairpin motif is involved in DNA binding (By
CC similarity).
CC -!- SIMILARITY: Belongs to the uvrB family.
CC -!- SIMILARITY: Contains 1 UVR domain.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; AB013589; AM32985.1; -.
CC HAMAP; MF_00204; -.
CC InterPro; IPR001410; DEAD.
CC InterPro; IPR001650; Helicase_C.
CC InterPro; IPR004807; UvrB.
CC InterPro; IPR001943; UvrB/C.
CC Pfam; PF00271; helicase_C; 1.
CC Pfam; PF02151; UVR; 1.
CC TIGRFAMs; TIGR00631; uvrB; 1.
CC PROSITE; PS0151; UVR; 1.
CC SOS response; Excision nuclease; DNA repair; DNA recombination;
CC DNA excision; ATP-binding; Complete proteome.
CC NP BIND 64 71
CC FT DOMAIN 117 140 BETA-HAIRPIN.
CC FT DOMAIN 531 566
CC SEQUENCE 670 AA; 77690 MW; C53D15B2210746EA CRC64;

Query Match 1.3%; Score 7; DB 1; Length 670;
Best Local Similarity 100.0%; Pred. No. 1.5e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 82 DGAQGE 88

RESULT 87
ID CA39 CHICK STANDARD; PRT; 675 AA.
AC P32017;
DT 01-JUL-1993 (Rel. 26, Created)
DT 01-JUL-1993 (Rel. 26, Last sequence update)
DT 15-MAR-2004 (Rel. 43, Last annotation update)
DE Collagen alpha 3(IX) chain precursor.
GN COL9A3.
OS Gallus gallus (Chicken).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;
OC Gallus.
OX NCBI_TaxID=9031;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=92250566; PubMed=1577778;
RA Hay-El R., Sharma Y.D., Aguilera A., Ueyama N., Wu J.J.,
RA Eyre D.R., Juricic L., Chandrasekaran S., Li M., Nah H.D.,
RA Upholt W.B., Tanzer M.L.;
RT "Cloning and developmental expression of the alpha 3 chain of chicken
RT type IX collagen.";
RL J. Biol. Chem. 267:10070-10076(1992).
CC -!- FUNCTION: COLLAGEN TYPE IX IS A MINOR CARTILAGE NONFIBRILLAR
CC COLLAGEN. IT IS ASSOCIATED WITH TYPE II COLLAGEN FIBRILS.
CC -!- SUBUNIT: TRIMERS COMPOSED OF THREE DIFFERENT CHAINS: ALPHA 1(IX),
CC ALPHA 2(IX), AND ALPHA 3(IX).
CC -!- PTM: Prolines at the third position of the tripeptide repeating
CC unit (G-X-Y) are hydroxylated in some or all of the chains.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; M83179; AB59960.1; -.
CC InterPro; IPR008161; C1g_helix.
CC InterPro; IPR008160; Collagen.
CC Pfam; PF01391; Collagen; 11.
CC ProDom; PD000007; C1g_helix; 3.
CC Extracellular matrix; Connective tissue; Repeat; Hydroxylation;
CC Glycoprotein; Collagen; Signal.
CC SIGNAL 1 21
CC CHAIN 22 675 COLLAGEN ALPHA 3(IX) CHAIN.
CC FT DOMAIN 25 515 NONHELICAL REGION 3 (COL3).
CC FT DOMAIN 516 546 NONHELICAL REGION 3 (NC3).
CC FT DOMAIN 547 626 TRIPLE-HELICAL REGION 2 (COL2).
CC FT DOMAIN 627 631 NONHELICAL REGION 2 (NC2).
CC FT DOMAIN 632 658 TRIPLE-HELICAL REGION 1 (COL1).
CC FT DOMAIN 659 675 NONHELICAL REGION 1 (NC1).
CC FT SITE 242 244 CELL ATTACHMENT SITE (POTENTIAL).
CC FT SITE 591 593 CELL ATTACHMENT SITE (POTENTIAL).
CC FT CARBOHYD 479 479 N-LINKED (GLCNAC... ) (POTENTIAL).
CC SEQUENCE 675 AA; 63013 MW; C983FEC924A10098 CRC64;

Query Match 1.3%; Score 7; DB 1; Length 675;
Best Local Similarity 100.0%; Pred. No. 1.5e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 82 DGAQGE 88
```

```
Db 631 DGAQEP 637
|||||
RESULT 88
ID HS7E DROME STANDARD; PRT; 687 AA.
AC 29845;
DT 01-APR-1993 (Rel. 25, Created)
DT 01-APR-1993 (Rel. 25, Last sequence update)
DT 01-JUN-1994 (Rel. 29, Last annotation update)
DE Heat shock 70 kDa protein cognate 5.
GN HSC70 OR HSC70-5.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=93292982; PubMed=8514184;
RA Rubin D.M., Mehta A., Zhu J., Shoham S., Chen X.J., Wells Q.,
RA Palter K.B.;
RT "Genomic structure and sequence analysis of Drosophila melanogaster
RT HSC70 genes.";
RL Gene 128:155-163(1993).
CC -!- DEVELOPMENTAL STAGE: Heat shock cognate proteins are expressed
CC constitutively during normal development.
CC -!- SIMILARITY: Belongs to the heat shock protein 70 family.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
DR EMBL; L01502; AAA28628.1; -
DR PIR; JN0667; JN0667.
DR HSSP; P04475; 1DG4.
DR FLYBase; FBgn001220; Hsc70-5.
DR GO; GO:0003773; P:heat shock protein activity; NAS.
DR InterPro; IPR001023; Hsp70.
DR Pfam; PF00012; HSP70; 1.
DR PRINTS; PR00301; HEATSHOCK70.
DR PRODom; PD000089; Hsp70; 1.
DR PROSITE; PS00297; HSP70_1; 1.
DR PROSITE; PS00329; HSP70_2; 1.
DR PROSITE; PS01036; HSP70_3; 1.
DR ATP-binding; Heat shock; Multigene family.
DR XW SEQUENCE 687 AA; 74208 MW; 43F6BCA1D2052EFB CRC64;
SQ
Query Match 1.3%; Score 7; DB 1; Length 687;
Best Local Similarity 100.0%; Pred. No. 1.5e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY 222 RSKLESL 228
|||||
Db 348 RSKLESL 354
|||||
RESULT 89
SYM_METAC
ID SYM METAC STANDARD; PRT; 712 AA.
AC QSTIUS;
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Methionyl-tRNA synthetase (EC 6.1.1.10) (Methionine--tRNA ligase)
DE (MetrS).
OS Methanosarcina acetivorans.
OX NCBI_TaxID=2214;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN=C2A / ATCC 35395 / DSM 2834;
RX MEDLINE=21929760; PubMed=11932338;
RA Galagan J.E., Nusbaum C., Roy A., Endrizzi M.G., Macdonald P.,
RA Fitzhugh W., Calvo S., Engels R., Smirnov S., Ancoor D., Brown A.,
RA Allen N., Naylor J., Stange-Thomann N., DeArellano K., Johnson R.,
RA Linton L., McEwan P., McKernan K., Talamas J., Tirrell A., Ye W.,
RA Zimmer A., Barber R.D., Cann I., Graham D.E., Grahame D.A., Guss A.M.,
RA Hedderich R., Ingram-Smith C., Kuetner H.C., Krzycki J.A.,
RA Leigh J.A., Li W., Liu J., Mukhopadhyay B., Reeve J.N., Smith K.,
RA Springer T.A., Umayam L.A., White O., White R.H., de Macario E.C.,
RA Perry J.G., Jarell K.P., Jing H., Macario A.J.L., Paulsen I.,
RA Pritchett M., Sowers K.R., Swanson K.V., Zinder S.H., Lander E.,
RA Metcalf W.W., Birren B.;
RT "The genome of Methanosarcina acetivorans reveals extensive metabolic
RT and physiological diversity.";
RL Genome Res. 12:532-542(2002).
CC -!- FUNCTION: Is required not only for elongation of protein synthesis
CC but also for the initiation of all mRNA translation through
CC initiator tRNA(Met) aminoacylation (By similarity).
CC -!- CATALYTIC ACTIVITY: ATP + L-methionine + tRNA(Met) = AMP +
CC diphosphate + L-methionyl-tRNA(Met).
CC -!- COFACTOR: Binds 1 zinc ion per subunit (By similarity).
CC -!- SUBUNIT: Homodimer (By similarity).
CC -!- SUBCELLULAR LOCATION: Cytoplasmic.
CC -!- SIMILARITY: Belongs to class-I aminoacyl-tRNA synthetase family.
CC MetG subfamily 1.
CC -!- SIMILARITY: Contains 1 tRNA-binding domain.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
DR EMBL; AE011117; AA07394.1; -
DR HAMAP; MF 00098; fused; 1.
DR InterPro; IPR004455; MetG_Cterm.
DR InterPro; IPR008224; MetRS_dimerising.
DR InterPro; IPR008994; Nucleic acid OB.
DR InterPro; IPR002300; tRNA-synt_1a.
DR InterPro; IPR001412; tRNA-synt_1.
DR InterPro; IPR002304; tRNA-synt_met.
DR InterPro; IPR002547; tRNA_bind.
DR Pfam; PF00133; tRNA-synt_1; 1.
DR Pfam; PF01588; tRNA_bind; 1.
DR FIRSP; FIRSF001528; MetRS_dimerising; 1.
DR PRINTS; PR01041; TRNASYNTHMET.
DR TIGRFAMs; TIGR00398; metG; 1.
DR TIGRFAMs; TIGR00399; metG_C term; 1.
DR PROSITE; PS00386; TRED; 1.
DR PROSITE; PS00178; AA_TRNA_LIGASE_1; FALSE_NEG.
DR Aminoacyl-tRNA synthetase; Protein biosynthesis; Ligase; ATP-binding;
DR RNA-binding; tRNA-binding; Metal-binding; Zinc; Complete proteome.
DR SITE 20 30 "HIGH" REGION.
DR SITE 334 338 "RMSKS" REGION.
DR FT DOMAIN 610 712 tRNA-BINDING.
DR FT METAL 151 151 ZINC (BY SIMILARITY).
DR FT METAL 154 154 ZINC (BY SIMILARITY).
DR FT METAL 163 163 ZINC (BY SIMILARITY).
DR FT METAL 167 167 ZINC (BY SIMILARITY).
DR FT BINDING 337 337 ATP (BY SIMILARITY).
SQ SEQUENCE 712 AA; 79747 MW; ECA52862E91BD25B CRC64;
Query Match 1.3%; Score 7; DB 1; Length 712;
Best Local Similarity 100.0%; Pred. No. 1.6e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```

QY 115 GEKEPSK 121
DB 569 GEKEPSK 575

RESULT 90
SYM METMA STANDARD; PRT; 715 AA.
AC Q9Y222; Q8WZ57; Q96F56; Q9BS06;
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DE Methionyl-tRNA synthetase (EC 6.1.1.10) (Methionine--tRNA ligase)
DE (MethRS).
DE
GN METG OR MW0867.
OS Methanosarcina mazei (Methanosarcina frisia).
OC Archaea; Euryarchaeota; Methanomicrobia; Methanosarcinales;
OC Methanosarcinaceae; Methanosarcina.
OX NCBI_TaxID=2209;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Goel / G01 / ATCC BAA-199 / DSM 3647 / OCM 88;
RX MEDLINE=22120827; PubMed=12125824;
RA Deppenmeier U., Johann A., Hartoch T., Merkl R., Schmitz R.A.,
RA Martinez-Arias R., Henne A., Wierzer A., Baeumer S., Jacobi C.,
RA Brueggemann H., Lienard T., Christmann A., Boemcke M., Steckel S.,
RA Bhattacharyya A., Lykidis A., Overbeek R., Klenk H.-P., Gunsalus R.P.,
RA Fritz H.-J., Gottschalk G.;
RT "The genome of Methanosarcina mazei: evidence for lateral gene
RT transfer between Bacteria and Archaea";
RL J. Mol. Microbiol. Biotechnol. 4:453-461(2002).
CC -!- FUNCTION: Is required not only for elongation of protein synthesis
CC but also for the initiation of all mRNA translation through
CC initiator tRNA(fMet) aminoacylation (By similarity).
CC -!- CATALYTIC ACTIVITY: ATP + L-methionine + tRNA(Met) = AMP +
CC diphosphate + L-methionyl-tRNA(Met).
CC -!- COFACTOR: Binds 1 zinc ion per subunit (By similarity).
CC -!- SUBUNIT: Homodimer (By similarity).
CC -!- SUBCELLULAR LOCATION: Cytoplasmic.
CC -!- SIMILARITY: Belongs to class-I aminoacyl-tRNA synthetase family.
CC MetG subfamily 1.
CC -!- SIMILARITY: Contains 1 tRNA-binding domain.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC
DR EMBL; AE013312; AM30563.1; ALT_INIT.
DR HAMAP; MF_00098; fused; 1.
DR InterPro; IPR004495; MetG Cterm.
DR InterPro; IPR008224; MetRS dimerising.
DR InterPro; IPR008994; Nucleic acid OB.
DR InterPro; IPR002300; tRNA-synt_1a.
DR InterPro; IPR001412; tRNA-synt_1.
DR InterPro; IPR002304; tRNA-synt_met.
DR InterPro; IPR002547; tRNA_bind.
DR Pfam; PF00133; tRNA-synt_1; 1.
DR Pfam; PF01588; tRNA bind; 1.
DR PRSF; PRSF001528; MetRS dimerising; 1.
DR PRINTS; PR01041; TRNASYNTMET.
DR TIGRFAMs; TIGR00398; metG; 1.
DR TIGRFAMs; TIGR00399; metG_Cterm; 1.
DR PROSITE; PS00178; AA TRNA_LIGASE_I; FALSE_NEG.
DR PROSITE; PS50886; TRBD; 1.
KW Aminoacyl-tRNA synthetase; Metal-binding; Zinc; Complete proteome.
FT SITE 20 30 "HIGH" REGION
FT SITE 334 338 "WMSKS" REGION.

```

Query Match 1.3%; Score 7; DB 1; Length 715;
Best Local Similarity 100.0%; Pred. No. 1.6e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 115 GEKEPSK 121
DB 569 GEKEPSK 575

RESULT 91
MTOL_HUMAN STANDARD; PRT; 717 AA.
ID MTOL_HUMAN
AC Q9Y222; Q8WZ57; Q96F56; Q9BS06;
DT 30-MAY-2000 (Rel. 39, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE MTOL protein homolog (CGI-02).
DE MTOL.
GN
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A. (ISOFORM 5).
RX MEDLINE=2072150; PubMed=10810093;
RA Lai C.-H., Chou C.-Y., Chang L.-Y., Liu C.-S., Lin W.-C.;
RT "Identification of novel human genes evolutionarily conserved in
RT Caenorhabditis elegans by comparative proteomics";
RL Genome Res. 10:703-713(2000).
RN [2]
RP SEQUENCE FROM N.A. (ISOFORMS 1; 2; 3 AND 5).
RA Li R.H., Li X.M., Zhang X., Guan W.X.;
RL Submitted (NOV-2001) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A. (ISOFORMS 5 AND 6).
RC TISSUE=Lymph, and Placenta;
RX MEDLINE=22388257; PubMed=12477932;
RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G., Schuler G.D.,
RA Klausner R.D., Collins F.S., Wagner K.H., Schefer C.F., Bhat N.K.,
RA Altschul S.F., Zeeberg B., Buetow K.H., Wang J., Hsieh F.,
RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hong L.,
RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
RA Brownstein M.J., Udgin T.B., Toshiyuki S., Carninci P., Prange C.,
RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullany S.J.,
RA Bosak S.A., McKean P.J., McKernan K.D., Malek J.A., Gunaratne P.H.,
RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
RA Villalón D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA Fahey J., Helton E., Kettman M., Maman A., Rodrigues S., Sanchez A.,
RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,
RA Butterfield Y.S.N., Krzywinski M.I., Skalska U., Smallos D.E.,
RA Scherch A., Jones S.J.M., Marra M.A.;
RT "Generation and initial analysis of more than 15,000 full-length
RT human and mouse cDNA sequences";
RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
CC -!- ALTERNATIVE PRODUCTS.
CC Event=Alternative splicing; Named isoforms=5;
CC Name=3; Synonyms=4;
CC IsoId=Q9Y222-1; Sequence=Displayed;
CC Name=1;
CC IsoId=Q9Y222-2; Sequence=VSP_001749; VSP_001751;
CC Name=2;
CC IsoId=Q9Y222-3; Sequence=VSP_001750;

```

CC Name=5;
CC IsoId=Q9Y222-4; Sequence=VSP_001751;
CC Name=6;
CC IsoId=Q9Y222-5; Sequence=VSP_001748;
CC -!- SIMILARITY: Belongs to the gldA family.
CC -!- CAUTION: Ref.1 sequence differs from that shown due to a
CC frameshift in position 9.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; AF132937; AAD27712.1; ALT_FRAME.
CC EMBL; AF459110; AAL82394.1; -.
CC EMBL; AF459111; AAL82395.1; -.
CC EMBL; AF442963; AAL35894.1; -.
CC EMBL; AF319422; AAG42814.3; -.
CC EMBL; AF078986; AAL85491.1; -.
CC EMBL; AF078985; AAL85490.1; -.
CC EMBL; BC005808; AAK05808.1; -.
CC EMBL; BC011051; AAH11051.1; ALT_INIT.
CC Genew; HGNC:19261; MTO1.
CC InterPro; IPR001327; EAD_pyr_redox.
CC InterPro; IPR002218; GIDA.
CC InterPro; IPR004416; GIDA_sub.
CC InterPro; IPR001100; Pyr_redox.
CC Pfam; PF01134; GIDA; 1.
CC PRINTS; PR00368; FADPNR.
CC PROSITE; PS00411; PNRDRTASE1.
CC PRODOM; PRO03738; GIDA; 1.
CC PROSITE; PS01280; GIDA_1; 1.
CC PROSITE; PS01281; GIDA_2; 1.
CC Alternative splicing.
CC KW VARSPPLIC 1 74
CC Missing (in isoform 6).
CC /FTid=VSP_001748.
CC Missing (in isoform 1).
CC /FTid=VSP_001749.
CC Missing (in isoform 2).
CC /FTid=VSP_001750.
CC Missing (in isoform 1 and isoform 5).
CC /FTid=VSP_001751.
CC H -> O (IN REP. 1 AND 2;
CC AAL82394/AAL82395).
CC LA -> CT (IN REF. 2; AAL35894).
CC SEQUENCE 717 AA; 79963 MW; 78F84D8633BC0ED3 CRC64;
CC -----
CC Query Match 1.3%; Score 7; DB 1; Length 717;
CC Best Local Similarity 100.0%; Pred.No. 1.6e+02;
CC Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
CC -----
CC QY 288 ELAERLK 294
CC DB 602 ELAERLK 608
CC -----
CC RESULT 92
CC CPTM HUMAN STANDARD; PRT; 772 AA.
CC AC Q92523; Q13389; Q99655;
CC DT 01-NOV-1997 (Rel. 35, Created)
CC DT 01-NOV-1997 (Rel. 35, Last sequence update)
CC DT 15-MAR-2004 (Rel. 43, Last annotation update)
CC DE Carnitine O-palmitoyltransferase I, mitochondrial muscle isoform
CC DE (EC 2.3.1.21) (CPT I) (CPTI-M) (Carnitine palmitoyltransferase I like
CC DE protein).
CC GN CPTIIB.
CC OS Homo sapiens (Human).
CC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
CC -----
CC Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.
CC NCBI_TaxID=9606;
CC [1]
CC SEQUENCE FROM N.A.
CC TISSUE=Heart;
CC RX MEDLINE=96283810; PubMed=8679700;
CC RA Yamazaki N., Shinohara Y., Shima A., Yamanaka Y., Terada H.;
CC RT "Isolation and characterization of cDNA and genomic clones encoding
CC RL human muscle type carnitine palmitoyltransferase I.";
CC RN Biochim. Biophys. Acta 1307:157-161(1996).
CC [2]
CC SEQUENCE FROM N.A.
CC RA Adams M.D.;
CC RL Submitted (JUL-1996) to the EMBL/GenBank/DBJ databases.
CC [3]
CC SEQUENCE FROM N.A.
CC RX MEDLINE=98008026; PubMed=9344464;
CC RA Zhu H., Shi J., de Vries Y., Arvidson D.N., Cregg J.M.,
CC Woldesorgis G.;
CC RT "Functional studies of yeast-expressed human heart muscle carnitine
CC RL palmitoyltransferase I.";
CC RN Arch. Biochem. Biophys. 347:53-61(1997).
CC [4]
CC SEQUENCE FROM N.A.
CC TISSUE=Heart;
CC RX MEDLINE=97224516; PubMed=9070950;
CC RA Britton C.H., Mackey D.W., Esser V., Foster D.W., Burns D.K.,
CC Yarnall D.P., Froguel P., McGarry J.D.;
CC RT "Fine chromosome mapping of the genes for human liver and muscle
CC RL carnitine palmitoyltransferase I (CPT1A and CPT1B).";
CC RN Genomics 40:209-211(1997).
CC [5]
CC SEQUENCE FROM N.A.
CC TISSUE=Skeletal muscle;
CC RA van der Leij F.R., Takens J., van der Veen A.Y., Terpstra P.,
CC Ruijters J.R.G.;
CC RL Submitted (MAR-1997) to the EMBL/GenBank/DBJ databases.
CC [6]
CC SEQUENCE FROM N.A.
CC RX MEDLINE=97367931; PubMed=9224698;
CC RA Yamazaki N., Yamanaka Y., Hashimoto Y., Shinohara Y., Shima A.,
CC Terada H.;
CC RT "Structural features of the gene encoding human muscle type carnitine
CC RL palmitoyltransferase I.";
CC RN PNAS Lett. 409:401-406(1997).
CC -!- CATALYTIC ACTIVITY: Palmitoyl-CoA + L-carnitine = CoA + L-
CC palmitoylcarnitine.
CC -!- PATHWAY: Fatty acid beta-oxidation cycle.
CC -!- SUBCELLULAR LOCATION: Mitochondrial outer membrane.
CC -!- TISSUE SPECIFICITY: Strong expression in heart and skeletal
CC muscle. No expression in liver and kidney.
CC -!- SIMILARITY: Belongs to the carnitine/choline acetyltransferase
CC family.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; D87812; BAA13461.1; -.
CC EMBL; U62317; AAB03343.1; -.
CC EMBL; U66828; AAB40651.1; -.
CC EMBL; U62733; AAC51122.1; -.
CC EMBL; Y08682; CAA69938.1; -.
CC EMBL; Y08683; CAA69939.1; -.
CC EMBL; AS003286; BAA21492.1; -.
CC PIR; G02860; G02860.
CC Genew; HGNC:2329; CPT1B.
CC MIM; 601987; -.
CC GO; GO:0005739; C-mitochondrion; TAS.

```

DR GO: 0004095; F:carnitine O-palmitoyltransferase activity; TAS.
 DR GO: 000635; P:fatty acid beta-oxidation; TAS.
 DR InterPro: IPR000542; Carn acyl trans.
 DR Pfam: PF00755; Carn acyltransf_1.
 DR PROSITE: PS00439; ACYLTRANSF_C_1; 1.
 DR PROSITE: PS00440; ACYLTRANSF_C_2; 1.
 KW Fatty acid metabolism; Mitochondrion; Outer membrane;
 KW Fatty acid metabolism; Transport; Transmembrane; Multigene family;
 KW Polymorphism.
 FT DOMAIN 1 47 CYTOPLASMIC (POTENTIAL).
 FT TRANSMEM 48 73 POTENTIAL.
 FT DOMAIN 74 102 MITOCHONDRIAL INTERMEMBRANE (POTENTIAL).
 FT TRANSMEM 103 122 POTENTIAL.
 FT DOMAIN 123 122 CYTOPLASMIC (POTENTIAL).
 FT ACT_SITE 473 473 POTENTIAL.
 FT VARIANT 531 531 E -> K (in dbSNP:470117).
 FT VARIANT 664 664 /FTID=VAR_011739.
 FT VARIANT 664 664 /FTID=VAR_011740.
 SQ SEQUENCE 772 AA; 87801 MW; P7E3ED40643DFC81 CRC64;
 Query Match 1.3%; Score 7; DB 1; Length 772;
 Best Local Similarity 100.0%; Pred. No. 1.7e+02;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 170 BEKLAAL 176
 DB 378 BEKLAAL 384
 RESULT 93
 CPTM RAT STANDARD; PRT; 772 AA.
 AC Q63704;
 DT 01-NOV-1997 (Rel. 35, Created)
 DT 01-NOV-1997 (Rel. 35, Last sequence update)
 DT 15-MAR-2004 (Rel. 43, Last annotation update)
 DE Carnitine O-palmitoyltransferase I, mitochondrial muscle isoform
 DE (BC 2.3.1.21) (CPT I) (CPTI-M) (Carnitine palmitoyltransferase I
 DE like protein).
 GN CPTIB.
 OS Rattus norvegicus (Rat).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
 OX NCBI_TaxID=10116;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=Wistar; TISSUE=Brown adipose tissue;
 RX MEDLINE=9246880; PubMed=7729550;
 RA Yamazaki N., Shinohara Y., Shima A., Terada H.;
 RT "High expression of a novel carnitine palmitoyltransferase I like
 RT protein in rat brown adipose tissue and heart: isolation and
 RT characterization of its cDNA clone.";
 RL FEBS Lett. 363:41-45(1995).
 RN [2]
 RP SEQUENCE FROM N.A., AND CHARACTERIZATION.
 RC STRAIN=Sprague-Dawley; TISSUE=Heart;
 RX MEDLINE=96215127; PubMed=8636126;
 RA Esser V., Brown N.F., Cowan A.T., Foster D.W., McGarry J.D.;
 RT "Expression of a cDNA isolated from rat brown adipose tissue and
 RT heart identifies the product as the muscle isoform of carnitine
 RT palmitoyltransferase I (M-CPT I). M-CPT I is the predominant CPT I
 RT isoform expressed in both white (epididymal) and brown adipocytes.";
 RL J. Biol. Chem. 271:6972-6977(1996).
 CC -1- CATALYTIC ACTIVITY: Palmitoyl-CoA + L-carnitine = CoA + L-
 CC palmitoylcarnitine.
 CC -1- PATHWAY: Fatty acid beta-oxidation cycle.
 CC -1- SUBCELLULAR LOCATION: Mitochondrial outer membrane.
 CC -1- TISSUE SPECIFICITY: High expression in heart, skeletal muscle and
 CC brown adipose tissue. Also expressed in white adipose tissue, but
 CC not in liver.
 CC -1- SIMILARITY: Belongs to the carnitine/choline acetyltransferase
 CC family.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR BMBL: D43623; BAA07733.1; -.
 DR PIR: S65532; S65532.
 DR InterPro: IPR000542; Carn acyl trans.
 DR Pfam: PF00755; Carn acyltransf_1.
 DR PROSITE: PS00439; ACYLTRANSF_C_1; 1.
 DR PROSITE: PS00440; ACYLTRANSF_C_2; 1.
 KW Transferase; Acyltransferase; Mitochondrion; Outer membrane;
 KW Fatty acid metabolism; Transport; Transmembrane; Multigene family.
 FT DOMAIN 1 47 CYTOPLASMIC (POTENTIAL).
 FT TRANSMEM 48 73 POTENTIAL.
 FT DOMAIN 74 102 MITOCHONDRIAL INTERMEMBRANE (POTENTIAL).
 FT TRANSMEM 103 122 POTENTIAL.
 FT DOMAIN 123 122 CYTOPLASMIC (POTENTIAL).
 FT ACT_SITE 473 473 POTENTIAL.
 SQ SEQUENCE 772 AA; 88216 MW; AFBB4BC67B47FA47 CRC64;
 Query Match 1.3%; Score 7; DB 1; Length 772;
 Best Local Similarity 100.0%; Pred. No. 1.7e+02;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 170 BEKLAAL 176
 DB 378 BEKLAAL 384
 RESULT 94
 CALD HUMAN STANDARD; PRT; 793 AA.
 ID CALD_HUMAN
 AC Q05682; Q13978; Q13979; Q14741; Q14742;
 DT 01-NOV-1997 (Rel. 35, Created)
 DT 01-NOV-1997 (Rel. 35, Last sequence update)
 DT 15-MAR-2004 (Rel. 43, Last annotation update)
 DE Caldesmon (CDM).
 GN CALDI OR CDM OR CAD.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A. (ISOFORM 4).
 RC TISSUE=Lung fibroblast;
 RX MEDLINE=91358497; PubMed=1885618;
 RA Novy R.B., Lin J.L.-C., Lin J.J.-C.;
 RT "Characterization of cDNA clones encoding a human fibroblast
 RT caldesmon isoform and analysis of caldesmon expression in normal and
 RT transformed cells.";
 RL J. Biol. Chem. 266:16917-16924(1991).
 RN [2]
 RP SEQUENCE FROM N.A. (ISOFORMS 1 AND 4).
 RC TISSUE=Aorta;
 RX MEDLINE=92209999; PubMed=1555769;
 RA Humphrey M.B., Herrera-Sosa H., Gonzalez G., Lee R., Bryan J.;
 RT "Cloning of cDNAs encoding human caldesmons.";
 RL Gene 112:197-204(1992).
 RN [3]
 RP SEQUENCE FROM N.A. (ISOFORMS 2; 3 AND 5).
 RX MEDLINE=93101679; PubMed=1465449;
 RA Hayashi K., Yano H., Hashida T., Takeuchi R., Takeda O., Asada K.,
 RA Takahashi B.-I., Kato I., Sobue K.;
 RT "Genomic structure of the human caldesmon gene.";
 RL Proc. Natl. Acad. Sci. U.S.A. 89:12122-12126(1992).
 RN [4]
 RP SEQUENCE FROM N.A. (ISOFORM 4).
 RC TISSUE=Skin;

RE MEDLINE-2238257; PubMed=12477932;
RA Klausner R.D., Feingold E.A., Grouse L.H., Derge J.C.,
RA Klausner R.D., Collins P.S., Wagner L., Shemen C.M., Schuler G.D.,
RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
RA Datchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
RA Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,
RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullaly S.J.,
RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA Richards S., Worley K.C., Hale S.E., Garcia A.M., Gay L.J., Hulyk S.W.,
RA Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA Fahy J., Helton E., Kettenan M., Madan A., Rodriguez S., Sanchez A.,
RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA Blakesley R.W., Touchman J.W., Green S.D., Dickson M.C.,
RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,
RA Butterfield Y.S.N., Krzywinski M.F., Skalska U., Smalley D.B.,
RA Scherch A., Schain J.E., Jones S.J.M., Marra M.A.,
RT "Generation and initial analysis of more than 15,000 full-length
RT human and mouse cDNA sequences.";
RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903 (2002).
CC -!- FUNCTION: Actin- and myosin-binding protein implicated in the
CC regulation of actomyosin interactions in smooth muscle and
CC nonmuscle cells (could act as a bridge between myosin and actin
CC filaments). Stimulates actin binding of tropomyosin which
CC increases the stabilization of actin filament structure. In muscle
CC tissues, inhibits the actomyosin ATPase by binding to p-actin.
CC This inhibition is attenuated by calcium-calmodulin and is
CC potentiated by tropomyosin. Interacts with actin, myosin, two
CC molecules of tropomyosin and with calmodulin. Also play an
CC essential role during cellular mitosis and receptor capping.
CC -!- SUBCELLULAR LOCATION: On thin filaments in smooth muscle and on
CC stress fibers in fibroblasts (nonmuscle) (By similarity).
CC -!- ALTERNATIVE PRODUCTS:
CC Event-Alternative splicing; Named isoforms=5;
CC Name=1; Synonyms=H-CAD;
CC IsoId=Q05682-1; Sequence=Displayed;
CC Name=2; Synonyms=WI-38 L-CAD I;
CC IsoId=Q05682-2; Sequence=VSP_004155;
CC Name=3; Synonyms=HELA L-CAD I;
CC IsoId=Q05682-3; Sequence=VSP_004154, VSP_004155;
CC Name=4; Synonyms=WI-38 L-CAD II, 1-CAD;
CC IsoId=Q05682-4; Sequence=VSP_004156;
CC Name=5; Synonyms=HELA L-CAD II;
CC IsoId=Q05682-5; Sequence=VSP_004154, VSP_004156;
CC -!- TISSUE SPECIFICITY: High-molecular-weight caldesmon (isoform 1) is
CC predominantly expressed in smooth muscles, whereas low-molecular-
CC weight caldesmon (isoforms 2, 3, 4 and 5) are widely distributed
CC in non-muscle tissues and cells. Not expressed in skeletal muscle
CC or heart.
CC -!- DOMAIN: The N-terminal part seems to be a myosin/calmodulin-
CC binding domain, and the C-terminal a tropomyosin/actin/calmodulin-
CC binding domain. These two domains are separated by a central
CC helical region in the smooth-muscle form.
CC -!- PTM: In non-muscle cells, phosphorylation by CDC2 during mitosis
CC causes caldesmon to dissociate from microfilaments.
CC Phosphorylation reduces caldesmon binding to actin, myosin, and
CC calmodulin as well as its inhibition of actomyosin ATPase
CC activity. Phosphorylation also occurs in both quiescent and
CC dividing smooth muscle cells with similar effects on the
CC interaction with actin and calmodulin and on microfilaments
CC reorganization (By similarity).
CC -!- SIMILARITY: Belongs to the caldesmon family.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See [http://www.isb-sib.ch/](http://www.isb-sib.ch/announce/)
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; M64110; AAA35636.1; -

DR EMBL; M83216; AAA58420.1; -
DR EMBL; M83216; AAA58419.1; -
DR EMBL; D90452; BAA14418.1; -
DR EMBL; D90453; BAA14419.1; -
DR EMBL; BC040354; AAB40354.1; -
DR PIR; JH0628; JH0628
DR Genew; HGNC:1441; CALD1.
DR GK; Q05682; -
DR MIN; 114213; -
DR GO; GO:0005856; C:cytoskeleton; TAS.
DR GO; GO:0003779; F:actin binding; TAS.
DR GO; GO:0005516; F:calmodulin binding; TAS.
DR GO; GO:0005523; F:tropomyosin binding; TAS.
DR InterPro; IPR006017; Caldesmon.
DR InterPro; IPR006018; Caldesmon_LSP.
DR Pfam; PF02029; Caldesmon; 1.
DR PRINTS; PR01076; CALDESMON.
DR KW Muscle protein; Actin-binding; Calmodulin-binding; Phosphorylation;
KW Repeat; Alternative splicing.
FT DOMAIN 319 375 3 X 14 AA TANDEM REPEATS OF B-E-E-K-R-A-
FT A-E-E-R-Q-R-I-K.
FT REPEAT 319 332 1.
FT REPEAT 333 346 2.
FT REPEAT 347 360 3.
FT DOMAIN 26 207 MYOSIN AND CALMODULIN-BINDING
FT (BY SIMILARITY).
FT DOMAIN 564 621 TROPOMYOSIN-BINDING (POTENTIAL).
FT DOMAIN 664 674 TROPOMYOSIN-BINDING (POTENTIAL).
FT DOMAIN 653 686 STRONG ACTIN-BINDING (BY SIMILARITY).
FT DOMAIN 716 722 CALMODULIN-BINDING (BY SIMILARITY).
FT DOMAIN 768 793 WEAK ACTIN-BINDING (BY SIMILARITY).
FT DOMAIN 39 46 POLY-ARG.
FT DOMAIN 81 86 POLY-THR.
FT DOMAIN 189 196 POLY-GLU.
FT DOMAIN 376 379 POLY-GLU.
FT DOMAIN 540 543 POLY-ARG.
FT DOMAIN 580 583 POLY-GLU.
FT DOMAIN 597 600 POLY-GLU.
FT MOD_RES 724 724 PHOSPHORYLATION (BY CDC2)
FT (BY SIMILARITY).
FT MOD_RES 730 730 PHOSPHORYLATION (BY CDC2)
FT (BY SIMILARITY).
FT MOD_RES 753 753 PHOSPHORYLATION (BY CDC2)
FT (BY SIMILARITY).
FT MOD_RES 759 759 PHOSPHORYLATION (BY CDC2)
FT (BY SIMILARITY).
FT MOD_RES 789 789 PHOSPHORYLATION (BY CDC2)
FT (BY SIMILARITY).
FT VARSPLIC 1 24 MODPERRERERKREMELEER -> MLGGSGSHGRSL
FT AALSQ (in isoform 3 and isoform 5).
FT VARSPLIC 208 436 /FTID=VSP_004154.
FT VARSPLIC 208 462 Missing (in isoform 2 and isoform 3).
FT VARSPLIC 208 462 /FTID=VSP_004155.
FT CONFLICT 530 530 Missing (in isoform 4 and isoform 5).
FT CONFLICT 530 530 V -> M (IN REF. 1).
FT SEQUENCE 793 AA; 93250 MW; 2A0DC63D16DD6B5F CRC64;
Query Match 1.3%; Score 7; DB 1; Length 793;
Best Local Similarity 100.0%; Pred. No. 1.7e-02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 248 REEEKR 254
Db 596 REEEKR 602
RESULT 95
SYQ_LUPLU STANDARD; PRT; 794 AA.
ID SYQ_LUPLU
AC P52780;
DT 01-OCT-1996 (Rel. 34, Created)
DT 15-DEC-1998 (Rel. 37, Last sequence update)

DT 15-DEC-1998 (Rel. 37, Last annotation update)
DE Glutaminyl-tRNA synthetase (EC 6.1.1.18) (Glutamine--tRNA ligase)
DE (GlnRS).
OS Lapisinus luteus (Yellow lupine).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids;
OC eurosids I; Fabales; Fabaceae; Papilionoideae; Genisteae; Lupinus.
OX NCBI_TaxID=3873;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=cv. Ventus.
RA Stasicka M., Rozek M., Barciszewski J.;
RT Isolation and characterization of a cDNA clone encoding a plant gene
RT of aminoacyl-tRNA synthetase";
RL (In) Plant Gene Registry: PGR35-103.
CC -1- CATALYTIC ACTIVITY: ATP + L-glutamine + tRNA (Gln) = AMP +
CC diphosphate + L-glutamyl-tRNA (Gln).
CC -1- SIMILARITY: Belongs to class-I aminoacyl-tRNA synthetase family.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; X91787; CAA62901.1; -
DR PIR; T09643; T09643.
DR HSSP; P00962; IGTR.
DR InterPro; IPR004514; Glns.
DR InterPro; IPR000924; Glu tRNA-synt 1c.
DR InterPro; IPR001412; tRNA-synt I.
DR InterPro; IPR007639; tRNA-synt 1c R1.
DR InterPro; IPR007638; tRNA-synt 1c R2.
DR Pfam; PF00749; tRNA-synt 1c; 1.
DR Pfam; PF03950; tRNA-synt 1c; 1.
DR Pfam; PF04558; tRNA-synt 1c R1; 1.
DR Pfam; PF04557; tRNA-synt 1c R2; 1.
DR PRINTS; PR00987; TRNASYNTHGLU.
DR TIGRFAMS; TIGR00440; glns; 1.
DR PROSITE; PS00178; AA TENA LIGASE I; 1.
KW Aminoacyl-tRNA synthetase; Protein biosynthesis; Ligase; ATP-binding.
FT SITE 277 287 "HIGH" REGION.
FT SITE 505 509 "KMSKS" REGION.
FT BINDING 508 508 ATP (BY SIMILARITY).
SQ SEQUENCE 794 AA; 90482 MW; 4C7D02C0795269E1 CRC64;

Query Match 1.3%; Score 7; DB 1; Length 794;
Best Local Similarity 100.0%; Pred. No. 1.7e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 319 QLVDAKL 325
DB 175 QLVDAKL 181

RESULT 96
Y104 SYNY3
ID Y104 SYNY3 STANDARD; PRT; 811 AA.
AC P54371;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Hypothetical protein slr0104.
GN SLR0104.
OS Synechocystis sp. (strain PCC 6803).
OC Bacteria; Cyanobacteria; Chroococcales; Synechocystis.
OX NCBI_TaxID=1148;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=96127529; PubMed=8590279;
RA Kaneke T., Tanaka A., Sato S., Kotani H., Sazuka T., Miyajima N.,

RA Sugita M., Tabata S.;
RT "Sequence analysis of the genome of the unicellular cyanobacterium
RT Synechocystis sp. strain PCC6803. I. Sequence features in the 1 Mb
RT region from map positions 64% to 92% of the genome.";
RL DNA Res. 2:153-166(1995).
CC -1- SUBCELLULAR LOCATION: Integral membrane protein (Potential).
CC -1- SIMILARITY: TO B SUBTILIS YQFF.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; D64004; BAAL0634.1; -
DR PIR; S76690; S76690.
DR InterPro; IPR006674; HD.
DR InterPro; IPR003607; Met phosphohydro.
DR InterPro; IPR006675; Unchar_HDIG.
DR Pfam; PF01966; HD; 1.
DR SMART; SM00471; HDC; 1.
DR TIGRFAMS; TIGR00277; HDIG; 1.
KW Hypothetical protein; Transmembrane; Complete proteome.
FT TRANSMEM 50 70 POTENTIAL.
FT TRANSMEM 376 396 POTENTIAL.
FT TRANSMEM 409 429 POTENTIAL.
FT TRANSMEM 442 462 POTENTIAL.
FT TRANSMEM 486 496 POTENTIAL.
FT TRANSMEM 496 516 POTENTIAL.
FT TRANSMEM 529 549 POTENTIAL.
SQ SEQUENCE 811 AA; 90254 MW; F6015BE8F804F255 CRC64;

Query Match 1.3%; Score 7; DB 1; Length 811;
Best Local Similarity 100.0%; Pred. No. 1.8e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 251 EKRKEV 257
DB 103 EKRKEV 109

RESULT 97
ENV HVIND
ID ENV HVIND STANDARD; PRT; 846 AA.
AC P18799;
DT 01-NOV-1990 (Rel. 16, Created)
DT 01-NOV-1990 (Rel. 16, Last sequence update)
DT 15-JUL-1999 (Rel. 38, Last annotation update)
DE Envelope glycoprotein GP160 precursor [Contains: Exterior membrane
DE glycoprotein (GP120); transmembrane glycoprotein (GP41)].
GN ENV.
OS Human immunodeficiency virus type 1 (NDK isolate) (HIV-1).
OC Viruses; Retroviridae; Retroviridae; Lentivirus.
OX NCBI_TaxID=11695;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=90034200; PubMed=2806917;
RA Spire B., Sire J., Zachar V., Rey F., Barre-Sinoussi F., Galibert F.,
RA Hampe A., Chermann J.C.;
RT "Nucleotide sequence of HIV-1-NDK: a highly cytopathic strain of the
RT human immunodeficiency virus.";
RL Gene 81:275-284(1989).
CC -1- MISCELLANEOUS: NDK, ISOLATED FROM A ZAIRIAN PATIENT AFFECTED WITH
CC AIDS, AND IS A HIGHLY CYTOPATHOGENIC STRAIN.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC

or send an email to license@isb-sib.ch).

```

CC EMBL; M27323; AAA44873.1; -.
DR PIR; JQ0066; VCLJND.
DR HIV; M27323; ENV5NDK.
DR InterPro; IPR000328; Env GP41.
DR InterPro; IPR000777; GP120.
DR Pfam; PF00516; GP120; 1.
DR Pfam; PF00517; GP41; 1.
KW AIDS; Coat protein; Polyprotein; Glycoprotein; Transmembrane;
KW Signal.
FT SIGNAL 1 29
FT CHAIN 30 501
FT CHAIN 502 846
FT DISULFID 53 73
FT DISULFID 118 200
FT DISULFID 125 191
FT DISULFID 130 152
FT DISULFID 213 242
FT DISULFID 223 234
FT DISULFID 291 328
FT DISULFID 374 435
FT DISULFID 381 408
FT CARBOHYD 87 87
FT CARBOHYD 129 129
FT CARBOHYD 151 151
FT CARBOHYD 179 179
FT CARBOHYD 182 182
FT CARBOHYD 229 229
FT CARBOHYD 236 236
FT CARBOHYD 257 257
FT CARBOHYD 271 271
FT CARBOHYD 284 284
FT CARBOHYD 290 290
FT CARBOHYD 351 351
FT CARBOHYD 382 382
FT CARBOHYD 388 388
FT CARBOHYD 392 392
FT CARBOHYD 395 395
FT CARBOHYD 401 401
FT CARBOHYD 438 438
FT CARBOHYD 451 451
FT CARBOHYD 452 452
FT CARBOHYD 501 501
FT CARBOHYD 506 506
FT CARBOHYD 515 515
FT CARBOHYD 527 527
SQ SEQUENCE 846 AA; 96476 MW; 8A3B9DA527DB2E83 CRC64;

Query Match 1.3%; Score 7; DB 1; Length 846;
Best Local Similarity 100.0%; Pred. No. 1.8e+02; Mismatches 0; Indels 0; Gaps 0;
Matches 7; Conservative 0;

QY 249 EEEEXRK 255
Db 144 EEEEXRK 150

RESULT 98
ARPB YEAST
ID ARPB YEAST STANDARD; PRT; 881 AA.
AC Q12386;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 15-JUL-1998 (Rel. 35, Last annotation update)
DE Actin-like protein ARP8.
GN ARP8 OR YOR141C OR YOR3348C.
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
OX NCBI_TaxID=4932;
RN [1]
RP SEQUENCE FROM N.A.

Query Match 1.3%; Score 7; DB 1; Length 846;
Best Local Similarity 100.0%; Pred. No. 1.8e+02; Mismatches 0; Indels 0; Gaps 0;
Matches 7; Conservative 0;

QY 249 EEEEXRK 255
Db 144 EEEEXRK 150

RESULT 98
ARPB YEAST
ID ARPB YEAST STANDARD; PRT; 881 AA.
AC Q12386;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 15-JUL-1998 (Rel. 35, Last annotation update)
DE Actin-like protein ARP8.
GN ARP8 OR YOR141C OR YOR3348C.
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
OX NCBI_TaxID=4932;
RN [1]
RP SEQUENCE FROM N.A.

or send an email to license@isb-sib.ch).
MEDLINE=97344368; PubMed=9200815;
Voss H., Benes V., Andrade M.A., Valencia A., Rechmann S., Teodoru C.,
Schwager C., Paces V., Sander C., Ansorge W.;
"DNA sequencing and analysis of 130 kb from yeast chromosome XV.";
Yeast 13:655-672(1997).
[2]
RP GENE NAME.
RX MEDLINE=97435478; PubMed=9290209;
Poch O., Winsor B.;
"Who's who among the Saccharomyces cerevisiae actin-related proteins?
A classification and nomenclature proposal for a large family.";
Yeast 13:1053-1058(1997).
CC -!- SIMILARITY: Belongs to the actin family.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
CC
DR EMBL; X94335; CAA64058.1; -.
DR EMBL; Z75049; CAA99341.1; -.
DR PIR; S67026; S67026.
DR GERMOnline; 143729; -.
DR SGD; S0005667; ARP8.
DR GO; GO:0005634; Cinnuleus; IDA.
DR InterPro; IPR004000; Actin_like.
DR Pfam; PF00022; actin; 1.
DR SMART; SM00268; ACTIN; 1.
DR KW Structural protein; Cytoskeleton.
FT DOMAIN 22 27
FT POLY-ASP.
SQ SEQUENCE 881 AA; 100208 MW; 8174851B6B077A19 CRC64;

Query Match 1.3%; Score 7; DB 1; Length 881;
Best Local Similarity 100.0%; Pred. No. 1.9e+02; Mismatches 0; Indels 0; Gaps 0;
Matches 7; Conservative 0;

QY 289 LAERLXX 295
Db 561 LAERLXX 567

RESULT 99
KLP5 SCHPO
ID KLP5 SCHPO STANDARD; PRT; 883 AA.
AC Q14343;
DT 15-JUL-1998 (Rel. 36, Created)
DT 15-JUL-1998 (Rel. 36, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Kinesin-like protein 5.
GN KLP5 OR SPBC2F12.13.
OS Schizosaccharomyces pombe (Fission yeast).
OC Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;
OC Schizosaccharomycetales; Schizosaccharomycetaceae;
OC Schizosaccharomycetes.
OX NCBI_TaxID=4896;
RN [1]
RP SEQUENCE FROM N.A., FUNCTION, SUBUNIT, AND SUBCELLULAR LOCATION.
RX MEDLINE=21963728; PubMed=11967147;
Garcia M.A., Koonruga N., Toda T.;
"Two kinesin-like Kin I family proteins in fission yeast regulate the
establishment of metaphase and the onset of anaphase A.";
Curr. Biol. 12:610-621(2002).
[2]
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=972;
RX MEDLINE=21848401; PubMed=11859360;
Wood V., Gwilliam R., Rajandream M.A., Lyne M., Lyne R., Stewart A.,
Scouras J., Peat N., Hayles J., Baker S., Basham D., Bowman S.,
Brooks K., Brown D., Brown S., Chillingworth T., Churcher C.M.,
Collins M., Connor R., Cronin A., Davis P., Feltwell T., Fraser A.,

```

RA Gentles S., Goble A., Hamlin N., Harris D., Hidalgo J., Hodgson G.,
RA Holroyd S., Hornsby T., Howarth S., Huckle E.J., Hunt S., Jagels K.,
RA James K., Jones L., Jones M., Leather S., McDonald S., McLean J.,
RA Mooney P., Moule S., Mungall K., Murphy L., Niblett D., Odell C.,
RA Oliver K., O'Neill S., Pearson D., Quail M.A., Rabinowitz E.,
RA Rutherford K., Rutter S., Saunders D., Seeger K., Sharp S.,
RA Skelton J., Simmonds M., Squares R., Stevens S.,
RA Taylor K., Taylor R.G., Tivey A., Walsh S.V., Warren T., Whitehead S.,
RA Woodward J., Volckaert G., Aert R., Robben J., Grymonprez B.,
RA Weltjens J., Vanstreels E., Rieger M., Schaefer M., Mueller-Auer S.,
RA Gabel C., Fuchs M., Fritzc C., Holzer G., Moestl D., Hilbert H.,
RA Borzym K., Langer I., Beck A., Leinhardt R., Reinhardt R., Pohl T.M.,
RA Eger P., Zimmermann W., Wedler H., Wambutt R., Purnelle B.,
RA Coffeau A., Cadieu E., Dreano S., Gloux S., Lelaure V., Mottier S.,
RA Galibert F., Aves S.J., Xiang Z., Hunt C., Moore K., Hurst S.M.,
RA Lucas M., Rochet M., Gaillardin C., Tallada V.A., Garzon A., Thode G.,
RA Daga R.R., Cruzado L., Jimenez J., Sanchez M., del Rey F., Benito J.,
RA Dominguez A., Revuelta J.L., Moreno S., Armstrong J., Forsburg S.L.,
RA Cerrutti L., Lowe T., McCombie W.R., Paulsen I., Potashkin J.,
RA Shpakovski G.V., Ussery D., Barrell B.G., Nurse P.,
RT Nature 415:871-880(2002).
CC -!- FUNCTION: Has a role in establishing metaphase during mitosis.
CC Required for chromosome segregation where it generates tension
CC during kinetochore capturing.
CC -!- SUBUNIT: Heterodimer with klp6.
CC -!- SUBCELLULAR LOCATION: Cytoplasmic microtubules in interphase,
CC mitotic kinetochores in metaphase and spindle midzone in anaphase
CC and telophase.
CC -!- SIMILARITY: BELONGS TO THE KINESIN-LIKE PROTEIN FAMILY. KINESIN
CC II SUBFAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; AB072924; BAB69885.1; -;
CC EMBL; Z597211; CAB10160.1; -;
CC PIR; T40128; T40128.
CC HSP; P71119; 3KAR.
CC GeneDB Spombe; SPBC2P12.13; -;
CC InterPro; IPR001752; kinesin_motor.
CC Pfam; PF00225; kinesin; 1.
CC PRINTS; PR00380; KINESINHEAVY.
CC SMART; SM00129; KISC; 1.
CC PROSITE; PS00411; KINESIN MOTOR DOMAIN1; 1.
CC PROSITE; PS00667; KINESIN MOTOR DOMAIN2; 1.
CC Mitosis; Chromosome partition; Motor protein; Microtubule;
CC ATP-binding; Coiled coil.
CC FT DOMAIN 1 388 KINESIN-MOTOR (BY SIMILARITY).
CC 396 435 COILED COIL (POTENTIAL).
CC FT DOMAIN 563 588 COILED COIL (POTENTIAL).
CC NP BIND 144 151 ATP (POTENTIAL).
CC SQ SEQUENCE 883 AA; 99058 MW; BBA60E22DA3978 CRC64;
CC
CC Query Match 1.3%; Score 7; DB 1; Length 883;
CC Best Local Similarity 100.0%; Pred No. 1.9e+02;
CC Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
CC
CC QY 222 RSKLSL 228
CC |||||
CC Db 514 RSKLSL 520
CC
CC RESULT 100
CC ID_F2_CHICK
CC KF2 CHICK STANDARD; PRT; 906 AA.
CC AC P98150;
CC DT 01-OCT-1996 (Rel. 34, Created)

DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 15-MAR-2004 (Rel. 43, Last annotation update)
DE Nuclear factor NF-kappa-B p100 subunit [Contains: Nuclear factor NF-
DE kappa-B p52 subunit].
GN NFkB2.
OS Gallus gallus (Chicken).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Archosauria; Aves; Neognathae; Galliformes; Phasianinae;
OC Gallus.
OX NCBI_TaxID=9031;
RN [1] _
RP SEQUENCE FROM N.A.
RC TISSUE=fibroblast;
RX MEDLINE=94047379; PubMed=8230480;
RA Sif S., Gilmore T.D.;
RT "NF-kappa B p100 is one of the high-molecular-weight proteins
RT complexed with the v-Rel oncoprotein in transformed chicken spleen
RT cells";
RL J. Virol. 67:7612-7617(1993).
RN [2]
RP SEQUENCE FROM N.A.
RC TISSUE=Spleen;
RX MEDLINE=94171036; PubMed=7510259;
RA Ikeda T., Hirota Y., Onodera T.;
RT "Isolation of a cDNA encoding the chicken p50B/p97 (Lyt-10)
RT transcription factor";
RL Gene 138:193-196(1994).
CC -!- FUNCTION: P100 is the precursor of the p52 subunit of the nuclear
CC factor NF-kappa-B, which binds to the kappa-B consensus sequence
CC 5'-GGGNNYYCC-3', located in the enhancer region of genes involved
CC in immune response and acute phase reactions. The precursor
CC protein itself does not bind to DNA.
CC -!- SUBUNIT: Active NF-kappa-B is a heterodimer of an about 52 kDa
CC DNA-binding subunit and the weak DNA-binding subunit p65. Two
CC heterodimers might form a labile tetramer.
CC -!- SUBCELLULAR LOCATION: Nuclear, but also found in the cytoplasm in
CC an inactive form complexed to an inhibitor (I-kappa-B).
CC -!- DOMAIN: The C-terminus of p100 might be involved in cytoplasmic
CC retention, inhibition of DNA-binding by p52 homodimers, and/or
CC transcription activation (By similarity).
CC -!- MISCELLANEOUS: NF-kappa B p100 is one of the high-molecular-weight
CC proteins complexed with the v-rel oncoprotein in transformed
CC chicken spleen cells.
CC -!- SIMILARITY: Belongs to the Rel/Dorsal family.
CC -!- SIMILARITY: Contains 6 ANK repeats.
CC -!- SIMILARITY: Contains 1 death domain.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; U00111; AAA03717.1; -;
CC EMBL; D16367; BAA0368.1; -;
CC PIR; I50404; I50404.
CC HSP; Q00653; IAKO.
CC TRANSFAC; T01929; -;
CC InterPro; IPR002110; ANK.
CC InterPro; IPR00488; Death.
CC InterPro; IPR007110; Ig-like.
CC InterPro; IPR002909; IPT_TIG.
CC InterPro; IPR000451; NF_Rel_dor.
CC InterPro; IPR008967; p53-like.
CC Pfam; PF00023; ank; 6.
CC Pfam; PF00531; death; 1.
CC Pfam; PF00554; RHD; 1.
CC Pfam; PF01833; TIG; 1.
CC PRINTS; PR00057; NFKBTNSCPFCT.
CC SMART; SM00248; ANK; 6.
CC SMART; SM00005; DEATH; 1.

DR SMART; SM00429; IPT; 1.
 DR PROSITE; PSS0088; ANK_REPEAT; 5.
 DR PROSITE; PSS0297; ANK_REPEAT; 1.
 DR PROSITE; PSS0017; DEATH_DOMAIN; FALSE_NEG.
 DR PROSITE; PSS01204; REL_1; 1.
 DR PROSITE; PSS0254; REL_2; 1.
 KW DNA-binding; Transcription regulation; Activator; Nuclear protein;
 KW Repeat; ANK repeat; Phosphorylation.
 FT DOMAIN 37 342 REL-LIKE [RHD].
 FT REPEAT 472 501 ANK 1.
 FT REPEAT 511 540 ANK 2.
 FT REPEAT 544 573 ANK 3.
 FT REPEAT 582 611 ANK 4.
 FT REPEAT 616 646 ANK 5.
 FT REPEAT 650 679 ANK 6.
 FT DOMAIN 771 857 DEATH.
 FT DOMAIN 336 340 NUCLEAR LOCALIZATION SIGNAL (POTENTIAL).
 FT DOMAIN 349 396 GLY-RICH.
 FT CONFLICT 55 55 V -> G (IN REF. 2).
 FT CONFLICT 272 272 S -> F (IN REF. 2).
 FT CONFLICT 337 337 N -> K (IN REF. 2).
 FT CONFLICT 395 395 C -> G (IN REF. 2).
 FT CONFLICT 409 409 R -> SG (IN REF. 2).
 FT CONFLICT 430 430 R -> A (IN REF. 2).
 FT CONFLICT 533 533 E -> Q (IN REF. 2).
 FT CONFLICT 569 569 A -> G (IN REF. 2).
 FT CONFLICT 653 653 T -> A (IN REF. 2).
 FT CONFLICT 690 693 VRVP -> SEA (IN REF. 2).
 FT CONFLICT 736 736 L -> RC (IN REF. 2).
 FT CONFLICT 759 774 SPILSCPPPPSRNHL -> RPDTEPTTPRAGNV (IN REF. 2).
 FT CONFLICT 779 779 T -> S (IN REF. 2).
 FT CONFLICT 792 792 Y -> D (IN REF. 2).
 FT CONFLICT 816 816 I -> D (IN REF. 2).
 FT CONFLICT 821 823 ASA -> SVSL (IN REF. 2).
 FT CONFLICT 830 830 P -> A (IN REF. 2).
 FT CONFLICT 885 885 R -> A (IN REF. 2).
 SQ SEQUENCE 906 AA; 99667 MW; 0E0CE20DB4F30B62 CRC64;
 Query Match 1.3%; Score 7; DB 1; Length 906;
 Best Local Similarity 100.0%; Pred. No. 1.9e+02;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 Qy 229 CRELQKH 235
 Db 423 CRELQKH 429
 Search completed: June 7, 2004, 14:45:41
 Job time : 23 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OK protein - protein search, using sw model

Run on: June 7, 2004, 14:42:11 ; Search time 19 Seconds
(without alignments)
2683.237 Million cell updates/sec

Title: US-10-023-529-8

Perfect score: 530

Sequence: 1 KSSFGQPEAGPEGAQERPSO.....APSTEASGQGPQEPISARA 530

Scoring table: OLIGO

Gapop 60.0 , Gapext 60.0

Searched: 283366 seqs, 96191526 residues

Word size : 0

Total number of hits satisfying chosen parameters: 283366

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 300 summaries

Database :

PIR 78: *
1: pir1: *
2: pir2: *
3: pir3: *
4: pir4: *

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	17	3-2	676	2 JC7222	77K muscle-derived
2	9	1-7	594	2 S50611	hypothetical prote
3	9	1-7	1402	2 I46707	translation initia
4	8	1-5	162	2 T29728	hypothetical prote
5	8	1-5	211	2 T23554	hypothetical prote
6	8	1-5	224	2 T52358	translation elonga
7	8	1-5	229	2 S87236	probable membrane
8	8	1-5	280	2 S71315	deoxyguanosine kin
9	8	1-5	281	2 C88538	protein F58P6.1 (i
10	8	1-5	312	2 T08985	hypothetical prote
11	8	1-5	450	2 E71909	hypothetical prote
12	8	1-5	467	2 D88710	hypothetical prote
13	8	1-5	500	2 H97769	protein C43G2.1 (i
14	8	1-5	525	2 H75514	hypothetical prote
15	8	1-5	531	2 A55887	glutamyl-tRNA synt
16	8	1-5	535	2 C36811	caldesmon, non-mus
17	8	1-5	535	2 A37994	Rf1 protein - salm
18	8	1-5	568	2 E90364	hypothetical prote
19	8	1-5	594	2 S62141	transcription init
20	8	1-5	758	2 S65169	hypothetical prote
21	8	1-5	761	2 T32183	hypothetical prote
22	8	1-5	928	2 T52292	hypothetical prote
23	8	1-5	952	2 T52456	endopeptidase Clp
24	8	1-5	952	2 T42983	AtClpC - Arabidops
25	8	1-5	1100	2 AE3243	conjugal transfer
26	8	1-5	1101	2 T03419	traA protein - Agr
27	8	1-5	1902	2 C97702	cell surface antige
28	8	1-5	2052	2 T18519	myosin X - bovine
29	7	1-3	23	2 S45032	homeotic protein S

30	7	1.3	60	2	B91253	hypothetical prote
31	7	1.3	62	2	AG1905	hypothetical prote
32	7	1.3	75	2	AC1755	hypothetical prote
33	7	1.3	79	2	D75326	hypothetical prote
34	7	1.3	105	2	S78290	ribosomal protein
35	7	1.3	120	2	S10587	cystatin C - rat
36	7	1.3	121	2	T44473	conserved hypotet
37	7	1.3	127	2	S07085	cystatin C precurs
38	7	1.3	128	2	T44497	hypothetical prote
39	7	1.3	133	2	S77948	major allergen Par
40	7	1.3	136	2	B65240	hypothetical prote
41	7	1.3	136	2	D64964	hypothetical prote
42	7	1.3	138	2	S52933	major allergen Par
43	7	1.3	138	2	A90258	conserved hypotet
44	7	1.3	154	2	G98112	hypothetical prote
45	7	1.3	155	2	B95248	hypothetical prote
46	7	1.3	158	2	T49567	related to attachm
47	7	1.3	159	2	T32043	hypothetical prote
48	7	1.3	163	2	E71183	VPS29-like phospho
49	7	1.3	172	2	A69454	hypothetical prote
50	7	1.3	176	2	AE2913	conserved hypotet
51	7	1.3	177	2	T37212	hypothetical prote
52	7	1.3	179	2	C85165	hypothetical prote
53	7	1.3	180	2	B33445	H+-transporting tr
54	7	1.3	181	2	E70209	conserved hypotet
55	7	1.3	187	2	E95846	probable ABC trans
56	7	1.3	190	2	T45013	hypothetical prote
57	7	1.3	193	2	A81149	outer membrane lip
58	7	1.3	193	2	H81874	probable outer mem
59	7	1.3	195	1	SAVLH1	delta large antige
60	7	1.3	195	1	SAVLH1	delta large antige
61	7	1.3	195	1	SAVLDM	delta large antige
62	7	1.3	195	2	S53112	delta large antige
63	7	1.3	199	2	G75028	h+-transporting Ar
64	7	1.3	201	2	G95852	conserved hypotet
65	7	1.3	211	2	C69375	hypothetical prote
66	7	1.3	214	1	SAVLDM	delta large antige
67	7	1.3	214	2	A53175	delta large antige
68	7	1.3	214	2	JC1062	delta large antige
69	7	1.3	218	2	S73675	hypothetical prote
70	7	1.3	225	2	T05514	hypothetical prote
71	7	1.3	229	1	PQBPE2	antiterminator Q -
72	7	1.3	229	2	D90831	antiterminator pr
73	7	1.3	229	2	C90912	probable antitermi
74	7	1.3	229	2	G90972	antiterminator (im
75	7	1.3	229	2	G85688	probable antitermi
76	7	1.3	229	2	C85820	antiterminator (im
77	7	1.3	232	2	G90292	hypothetical prote
78	7	1.3	232	2	E97347	probable transcrip
79	7	1.3	235	2	A86150	TIN6.24 protein -
80	7	1.3	237	2	H97687	hypothetical 17.5K
81	7	1.3	237	2	H97687	hypothetical prote
82	7	1.3	239	2	B95374	hypothetical prote
83	7	1.3	240	2	F90885	hypothetical prote
84	7	1.3	240	2	A85733	hypothetical prote
85	7	1.3	240	2	S84897	hypothetical prote
86	7	1.3	246	2	H72732	hypothetical prote
87	7	1.3	248	1	JH0252	myelin P0 protein
88	7	1.3	253	2	G72598	probable ABC trans
89	7	1.3	262	2	T04809	hypothetical prote
90	7	1.3	271	2	T37222	hypothetical prote
91	7	1.3	277	2	J06142	probable secreted
92	7	1.3	280	2	AI0182	deoxyguanosine kin
93	7	1.3	285	2	C86423	carboxylesterase {
94	7	1.3	297	2	P64750	unknown protein, i
95	7	1.3	306	2	T21938	regulatory protein
96	7	1.3	308	2	G84701	hypothetical prote
97	7	1.3	319	2	D64303	hypothetical prote
98	7	1.3	322	2	T45568	hypothetical prote
99	7	1.3	326	1	E64547	nitrogen fixation
100	7	1.3	326	2	B71960	hypothetical prote
101	7	1.3	327	2	B82918	DNA-directed RNA p
102	7	1.3	332	2	C90278	hypothetical prote
						conserved hypotet

```

; APPLICANT: Millet, Isabelle
; APPLICANT: Padigaru, Muralidhara
; APPLICANT: Patturajan, Meera
; APPLICANT: Rastelli, Luca
; APPLICANT: MacDougall, John R.
; APPLICANT: Mishra, Vishnu
; APPLICANT: Pena, Carol B.A.
; APPLICANT: Spaderna, Steven K.
; APPLICANT: Shimkets, Richard A.
; APPLICANT: Smithson, Glenda
; APPLICANT: Spytek, Kimberly A.
; APPLICANT: Stone, David J.
; APPLICANT: Shenoy, Suresh G.
; APPLICANT: Ort, Tatiana
; APPLICANT: Taupier Jr, Raymond J.
; APPLICANT: Tchernev, Velizar T.
; APPLICANT: Vernet, Corine A.M.
; APPLICANT: Wolenc, Adam R.
; APPLICANT: Zernuoen, Bryan D.
; APPLICANT: Zhong, Mei
; TITLE OF INVENTION: NOVEL PROTEINS AND NUCLEIC ACIDS ENCODING SAME
; FILE REFERENCE: 21402-533C
; CURRENT APPLICATION NUMBER: US/10/336,472
; CURRENT FILING DATE: 2003-01-03
; PRIOR APPLICATION NUMBER: 09/746,491
; PRIOR FILING DATE: 2000-12-20
; PRIOR APPLICATION NUMBER: 10/005,041
; PRIOR FILING DATE: 2001-12-04
; PRIOR APPLICATION NUMBER: 10/023,681
; PRIOR FILING DATE: 2001-12-18
; PRIOR APPLICATION NUMBER: 10/024,212
; PRIOR FILING DATE: 2001-12-18
; PRIOR APPLICATION NUMBER: 10/055,569
; PRIOR FILING DATE: 2001-10-26
; PRIOR APPLICATION NUMBER: 10/080,334
; PRIOR FILING DATE: 2002-02-21
; PRIOR APPLICATION NUMBER: 10/092,900
; PRIOR FILING DATE: 2002-03-07
; PRIOR APPLICATION NUMBER: 10/136,826
; PRIOR FILING DATE: 2002-05-01
; PRIOR APPLICATION NUMBER: 10/236,417
; PRIOR FILING DATE: 2002-09-06
; PRIOR APPLICATION NUMBER: 60/345,092
; PRIOR FILING DATE: 2002-01-04
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 230
; SOFTWARE: CuraseqList version 0.1
; SEQ ID NO 16
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-336-472-16

Query Match 8.8%; Score 236.5; DB 12; Length 1137;
Best Local Similarity 24.6%; Pred. No. 6e-06;
Matches 147; Conservative 80; Mismatches 196; Indels 175; Gaps 26;

QY 12 EGAQERPSQAAPAYEAEGCGSSQAPRKPGEQAQARTAGSALRDVSELSQLEDILSTYC 71
DB 237 EGSLEQEKVEMDLE-----RAKKLEGDLKLTQSS--IMDLND-KQOLEERLKKD 286
QY 72 VDNQGGPGEQAGCEPAEPEDA-----EKSRTYVAR-----NGEPEPTPVYGEKEP 119
DB 287 FSLN-----ALNARIEDQALGSQLOKQKLELOARIBELESELSAERTAKRAVKLR 338
QY 120 SKGDPNTEIRQS-DEVG-----DRHRPQKKKAKGLGKEITLLMQLNTLSTPEE 171
DB 339 SLSRELEBEISERLEEAGGATSVQIEMNKKRAEFQWRDLSEATL-----CHEA 389
QY 172 KLAALCKKY-----BELSEHNSOKMKLQKQSQVLQVKDHLRGHSHKAVLARSKLES 227
DB 390 TAAALRKKHADSVAELGQIDNLRQVKLEKSEFPLELDVDTSNMEQIIKAKANLEK 449

; APPLICANT: Ghosh, Soumitra S.
; APPLICANT: Fahy, Eoin D.
; APPLICANT: Gibson, Bradford W.
; APPLICANT: Taylor, Steven W.
; APPLICANT: Glenn, Gary M.
; APPLICANT: Warnock, Dale B.
; TITLE OF INVENTION: TARGETS FOR THERAPEUTIC INTERVENTION
; TITLE OF INVENTION: IDENTIFIED IN THE MITOCHONDRIAL PROTEOME
; FILE REFERENCE: 660088.465
; CURRENT APPLICATION NUMBER: US/10/408,765A
; CURRENT FILING DATE: 2003-04-04
; NUMBER OF SEQ ID NOS: 3077
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 254
; LENGTH: 3225
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-408-765A-254

Query Match 8.8%; Score 236.5; DB 16; Length 3225;
Best Local Similarity 22.8%; Pred. No. 2.2e-05;
Matches 119; Conservative 99; Mismatches 178; Indels 125; Gaps 18;

QY 52 LRDVSELSQLEDILSTYCVDNQGGPGEQAGCEPAEPEDAERSRTYVARNGSPEPTP 111
DB 2248 LOGENKELLSQLESTRHLY-----HSSQELAKLESELKSL-----KDQLTD 2289
QY 112 VYGEKEPSGDPNTEIRQSDEVGDRDHRPQ-----KKAKGLGKEITLLMQLT 162
DB 2290 LNSLEKCKEKGKNGLEGIIRQOEADIQSKFSVEQLETLQASRELTSLRHEINKEQK 2349
QY 163 LNTLSTPEKLAALCKKVAELLEHNSSOKMKLQKQSQVLQVKDHLRGHSHKAVLAR 222
DB 2350 IISLSGKSE--AIQVAIAELRQOH---DKRIKLENLLSQEEENIVLEENKAV--- 2401
QY 223 SKLESICRELRHNSLKEEGVQRAFEERKEKVS-----HFQVT 264
DB 2402 DKTNQLMETL-----KTIKKENIQQAQLDSFVKSSSLQNDRDRIVDYQLEERHLSII 2457
QY 265 LNDIQLOMQHHERNSKLQKQ-----NMELAEKLKLIQYELRBEEDHDKVFKH 313
DB 2458 LEKDQLIQAAAB--NNKLEKRIRGLRSHMDLNSNAKLDAELI-QY-----REDLNQVITI 2512

228 LCR-----ELQHN-----RSLKE----- 241
450 MCKTLEDQWNEHRSKAEETQRSVNDLTSQRAKLOTENGELSRLQDEKEALISQLTRGKIT 509
242 -----EGVQRAFEERKEKVS-FQVTLNDIQLOMQHNER-----NSK 281
510 YTOQLEDUKRQLEEEVKAKNALAHALQARHDCDLLRQYEEETAKAELQRLVSKANSE 569
282 LR-----QENMELAERLKKLIQYELRESHIDKV-----FKHK---DLQO 318
570 VAQWRTKYETDAIORTLEELAEKKLAQKLOBAEEAVEAVNAKSCSLEKTKHRLQNEIED 629
319 QLYDAKLOQAQEMLEKAEERHOREKDFLLKNAVESORMCELMKQCEETHLQKQALYTRKF 378
630 LAMVDVERNSA-----AAAALDKKORNFDKILAE--WKQYEESSOSELESSOKERSLSTELF 684
379 EEFONTLSKSSSEVFTTFQKEMEKTKKI-----KKLEKETMYR 417
685 -KLKNAYEESLEHLETFKREKNKLOBEISDLTEQLGSSGKTHLELKVRLQLEAKQNELQ 743
418 SWHESNKLALMAEKKVVRDELEGLQVKIQRLEKLC-----RALQTERNDLNKRVQD 471
744 SALEEA-EASLEHHEGKILR-AQLEFNQIKABIERKLAEKDEMEQAKNEL--RVVD 797

RESULT 43
US-10-408-765A-254
; Sequence 254, Application US/10408765A
; Publication No. US20040101874A1
; GENERAL INFORMATION:
; APPLICANT: Ghosh, Soumitra S.
; APPLICANT: Fahy, Eoin D.
; APPLICANT: Gibson, Bradford W.
; APPLICANT: Taylor, Steven W.
; APPLICANT: Glenn, Gary M.
; APPLICANT: Warnock, Dale B.
; TITLE OF INVENTION: TARGETS FOR THERAPEUTIC INTERVENTION
; TITLE OF INVENTION: IDENTIFIED IN THE MITOCHONDRIAL PROTEOME
; FILE REFERENCE: 660088.465
; CURRENT APPLICATION NUMBER: US/10/408,765A
; CURRENT FILING DATE: 2003-04-04
; NUMBER OF SEQ ID NOS: 3077
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 254
; LENGTH: 3225
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-408-765A-254

```


Db 1492 YVESDLETLKRNKNLQEQISDLTEQIAGGKRIHELEKIKKQVEQEKSELQAALSEA 1551
Qy 445 QVKIORLEKLCALQTERNDLKNRV 469
Db 1552 EASLEHEEGKILRIQLELNQVAKSEV 1576

RESULT 46
US-10-104-047-3636
; Sequence 3636, Application US/10104047
; Publication No. US20030236392A1
; GENERAL INFORMATION:
; APPLICANT: Anderson, David W.
; APPLICANT: Ballinger, Robert A.
; APPLICANT: Baumgartner, Jason C.
; APPLICANT: Burgess, Catherine E.
; APPLICANT: Casman, Stacie J.
; APPLICANT: Chant, John S.
; APPLICANT: Berghs, Constance
; APPLICANT: Gangolli, Esha A.
; APPLICANT: Edinger, Shlomit R.
; APPLICANT: Ellerman, Karen
; APPLICANT: Furtak, Katarzyna
; APPLICANT: Gerlach, Valerie
; APPLICANT: Gilbert, Jennifer A.
; APPLICANT: Gunther, Erik
; APPLICANT: Gorman, Linda
; APPLICANT: Guo, Xiaojia Sasha
; APPLICANT: Ji, Weizhen
; APPLICANT: Li, Li
; APPLICANT: Liu, Xiaohong
; APPLICANT: Miller, Charles E.
; APPLICANT: Millet, Isabelle
; APPLICANT: Padigaru, Muralidhara
; APPLICANT: Pasturajan, Meera
; APPLICANT: Rastelli, Luca
; APPLICANT: Macdougall, John R.
; APPLICANT: Mishra, Vishnu
; APPLICANT: Pena, Carol E.A.
; APPLICANT: Spaderna, Steven K.
; APPLICANT: Shinkets, Richard A.
; APPLICANT: Smithson, Glennnda
; APPLICANT: Spytek, Kimberly A.
; APPLICANT: Stone, David J.
; APPLICANT: Shenoy, Suresh G.
; APPLICANT: Ort, Tatiana
; APPLICANT: Taupier Jr, Raymond J.
; APPLICANT: Tchernev, Velizar T.
; APPLICANT: Vernet, Corine A.M.
; APPLICANT: Wolenc, Adam R.
; APPLICANT: Zernusen, Bryan D.
; APPLICANT: Zhong, Mei
; TITLE OF INVENTION: NOVEL PROTEINS AND NUCLEIC ACIDS ENCODING SAME
; FILE REFERENCE: 21402-533C
; CURRENT APPLICATION NUMBER: US/10/336,472
; CURRENT FILING DATE: 2003-01-03
; PRIOR APPLICATION NUMBER: 09/746,491
; PRIOR FILING DATE: 2000-12-20
; PRIOR APPLICATION NUMBER: 10/005,041
; PRIOR FILING DATE: 2001-12-04
; PRIOR APPLICATION NUMBER: 10/023,681
; PRIOR FILING DATE: 2001-12-18
; PRIOR APPLICATION NUMBER: 10/024,212
; PRIOR FILING DATE: 2001-12-18
; PRIOR APPLICATION NUMBER: 10/055,569
; PRIOR FILING DATE: 2001-10-26
; PRIOR APPLICATION NUMBER: 10/080,334
; PRIOR FILING DATE: 2002-02-21
; PRIOR APPLICATION NUMBER: 10/092,900
; PRIOR FILING DATE: 2002-03-07
; PRIOR APPLICATION NUMBER: 10/136,826
; PRIOR FILING DATE: 2002-05-01
; PRIOR APPLICATION NUMBER: 10/236,417
; PRIOR FILING DATE: 2002-09-06
; PRIOR APPLICATION NUMBER: 60/345,092
; PRIOR FILING DATE: 2002-01-04
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 230
; SOFTWARE: CuraSeqList version 0.1
; SEQ ID NO 20

Query Match 8.7%; Score 234.5; DB 15; Length 650;
Best Local Similarity 21.6%; Pred. No. 4e-06;
Matches 126; Conservative 100; Mismatches 257; Indels 99; Gaps 19;

Qy 7 PEAGPGAGERPQQAFAVEAGPGSQAPRPEGAQARTAGSGLRDVSEB-----LS 60
Db 58 PETTSGGCHSP-----EUTQNRQLKEEKVASHQHQBALREIEAODHTIRILT 108
Qy 61 RQLEDILSTYCDVNNQGGPGDGAQGPAPPEDAEKSRITYVARNGBP---EPTPVYVYGRK 117
Db 109 CQKTELETALYYGDAARFEDGNLTGTFSSFNLSQ---AFPGSPGCVSTSLIPGSS 164
Qy 118 EPSGSDPN-----TEIRQSDVGDHRRPQKKVAKGLGKEITLLMQTL-----NTL 166
Db 165 KDLAGRLHSHWHPAGELQALSASVSTRH-----KKADRYIELTKERDALSLELYRNTI 218
Qy 167 STPEEKLALCKVYAELEERHNSQMKLLQKQSQLVQEKDHLRG--EHSKAVILARSK 224
Db 219 TNEELK-----KNREL-----QKPLAESEKSIQNLVKELKRLERAKFLPLPQVQ 266
Qy 225 LESLCRELQRHNSLKEEGVORAREBEKKKVEYTSHPQVTLNDIQLQMEQHNRNSKLQ 284
Db 267 TNLQEMMRQEBELREQ-EKKIRQBEKMRQEBERLEREQEGKMRQEEKMMRQEKRLRE 325
Qy 285 ENMELAEKLKLEQVRELREHIDKVPFKHDLQQLVDKLAQQAQELKAEERHOREKD 344
Db 326 QEKLEQEKELREKQKLRQ-----EQMORQ--EEKWQBEKMRQEEKMRQEBE 376
Qy 345 FLKAEVESQRMCELMKQEQTHLQKQALYATEKPEEP--QNTLSKSEVPTTPK----- 396
Db 377 RLWEQEQKMRQEQQMRQDQERNWQEDRLREKMRQEQKMRQEQKMRQEEKMRQEQE 436
Qy 397 -----QEKMTKKIKGL-EKETT-----YRSRWSSNK-----ALLENA 431
Db 437 KKTQDQEKQOEERIRERBKKMRQEBEETVREQEKQKQKQERNWQEEKQEQQLPEQ 496
Qy 432 EEXTVRDKELEGQVKIQRLEKLCRALQTERNDLNKRVQLSAGGQSGLTDSOPER---R 488
Db 497 KEKLWEQEKQOEKLEWEQEKIRDOEMWQEKQKQWQBEKQWQBEKMRREKIRDO 556
Qy 489 PEGGGAQAPSPPVTEAPCPGAPSTASQGTGQPEPTSARA 530
Db 557 KKKMQERLPEHERCEBPCPLPPSKVLCNMGHSITVSVEPAGGEA 598

RESULT 47
US-10-336-472-20
; Sequence 20, Application US/10336472
; Publication No. US20040043929A1
; GENERAL INFORMATION:
; APPLICANT: Anderson, David W.
; APPLICANT: Ballinger, Robert A.
; APPLICANT: Baumgartner, Jason C.
; APPLICANT: Burgess, Catherine E.
; APPLICANT: Casman, Stacie J.
; APPLICANT: Chant, John S.
; APPLICANT: Berghs, Constance
; APPLICANT: Gangolli, Esha A.
; APPLICANT: Edinger, Shlomit R.
; APPLICANT: Ellerman, Karen
; APPLICANT: Furtak, Katarzyna
; APPLICANT: Gerlach, Valerie
; APPLICANT: Gilbert, Jennifer A.
; APPLICANT: Gunther, Erik
; APPLICANT: Gorman, Linda
; APPLICANT: Guo, Xiaojia Sasha
; APPLICANT: Ji, Weizhen
; APPLICANT: Li, Li
; APPLICANT: Liu, Xiaohong
; APPLICANT: Miller, Charles E.
; APPLICANT: Millet, Isabelle
; APPLICANT: Padigaru, Muralidhara
; APPLICANT: Pasturajan, Meera
; APPLICANT: Rastelli, Luca
; APPLICANT: Macdougall, John R.
; APPLICANT: Mishra, Vishnu
; APPLICANT: Pena, Carol E.A.
; APPLICANT: Spaderna, Steven K.
; APPLICANT: Shinkets, Richard A.
; APPLICANT: Smithson, Glennnda
; APPLICANT: Spytek, Kimberly A.
; APPLICANT: Stone, David J.
; APPLICANT: Shenoy, Suresh G.
; APPLICANT: Ort, Tatiana
; APPLICANT: Taupier Jr, Raymond J.
; APPLICANT: Tchernev, Velizar T.
; APPLICANT: Vernet, Corine A.M.
; APPLICANT: Wolenc, Adam R.
; APPLICANT: Zernusen, Bryan D.
; APPLICANT: Zhong, Mei
; TITLE OF INVENTION: NOVEL PROTEINS AND NUCLEIC ACIDS ENCODING SAME
; FILE REFERENCE: 21402-533C
; CURRENT APPLICATION NUMBER: US/10/336,472
; CURRENT FILING DATE: 2003-01-03
; PRIOR APPLICATION NUMBER: 09/746,491
; PRIOR FILING DATE: 2000-12-20
; PRIOR APPLICATION NUMBER: 10/005,041
; PRIOR FILING DATE: 2001-12-04
; PRIOR APPLICATION NUMBER: 10/023,681
; PRIOR FILING DATE: 2001-12-18
; PRIOR APPLICATION NUMBER: 10/024,212
; PRIOR FILING DATE: 2001-12-18
; PRIOR APPLICATION NUMBER: 10/055,569
; PRIOR FILING DATE: 2001-10-26
; PRIOR APPLICATION NUMBER: 10/080,334
; PRIOR FILING DATE: 2002-02-21
; PRIOR APPLICATION NUMBER: 10/092,900
; PRIOR FILING DATE: 2002-03-07
; PRIOR APPLICATION NUMBER: 10/136,826
; PRIOR FILING DATE: 2002-05-01
; PRIOR APPLICATION NUMBER: 10/236,417
; PRIOR FILING DATE: 2002-09-06
; PRIOR APPLICATION NUMBER: 60/345,092
; PRIOR FILING DATE: 2002-01-04
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 230
; SOFTWARE: CuraSeqList version 0.1
; SEQ ID NO 20

```

; LENGTH: 1859
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-336-472-20

Query Match      8.7%; Score 234; DB 12; Length 1859;
Best Local Similarity 23.4%; Pred. No. 1.6e-05;
Matches 117; Conservative 101; Mismatches 175; Indels 108; Gaps 24;

QY 8 EAGPGGAQERPSQAAPAV-----EAGPGSSQAPRPEGAQARTAGSGLRDVSEELS 60
Db 1363 KANSEVAQWRKYETDAIQRTTELEEAQKLAQRLQBAEEAVEAVNAKSSLEKTKRLQ 1422
QY 61 RQLEDILSTYCVDNNGGPGDGAQEPAPEDAKSRITYVARNGEPTPTVYGEKPS 120
Db 1423 NEIEDLM-----VDVRSNAAAAAL-----DKQR 1447
QY 121 KGPDPNTEIROSDEVGDRDRHRRPOEKKKAGLGKBITLIM-----QTLNLTSTPBEKLAAL 176
Db 1448 NFDKILAEWKQYBESQSELESSQ--KEARSLSLSTELFKLKNAYBESLEHLETFKRENKNL 1505
QY 177 CKKYABLLSEHRNSQOMKLLQKQSOQLVQKDHLRG-----EHSKAVLARSKLE- 226
Db 1506 QBEISDLTEQLGSSGKTIHELEKVRKQLEAKEMQLQSALEBAASLEHEGKILRAQLEF 1565
QY 227 -SLCRELQHRNSLKEBVGVARAEEREEKKEVTSHPQVTINDIQLQME--CHNERNSKLRQ 284
Db 1566 NQIKABIER-----KLAEKDEMEQAKRNHLRV--VDSLQTSLDATETSREALRV 1614
QY 285 ENWELAEKLLKLEQVELRE--EHDIVKFKHKL---QQQLVDKLAQQAQEMKLEA--RE 337
Db 1615 KKKQ-TSRKLAQEQ-ELIETSERVQLHSQNTSLNKKNDADLSLQTFEVEAVQBC 1672
QY 338 RHQREKDFLLKAVESQRC--ELMKQOBT--HLKQALALYTEKPEFQNTLSKSSEV-F 392
Db 1673 RNABEK---AKRAITDAAMAELEKKEQTSALHREKQKMGQTTKOLQHRLDREAQIAL 1729
QY 393 TTFQKQEMRMTKKIKLEKETTMYRSRWBSNKALEMAEETKVRDKB-LEGLOVKIQRL 451
Db 1730 KGGKQKQKLEARVRELENE-----LE-AEQK--RNAESVKGMRKSERRI 1771
QY 452 EKLCRALQTERNDLNRKVODL 472
Db 1772 KELTYQTEEDRKNL-LRLQDL 1791

RESULT 48
US-10-336-472-22
; Sequence 22, Application US/10336472
; Publication No.: US20040043929A1
; GENERAL INFORMATION:
; APPLICANT: Anderson, David W.
; APPLICANT: Ballinger, Robert A.
; APPLICANT: Baumgartner, Jason C.
; APPLICANT: Burgess, Catherine E.
; APPLICANT: Casman, Stacie J.
; APPLICANT: Chant, John S.
; APPLICANT: Berghs, Constance
; APPLICANT: Gangolli, Esha A.
; APPLICANT: Edinger, Shlomit R.
; APPLICANT: Ellerman, Karen
; APPLICANT: Furtak, Katarzyna
; APPLICANT: Gerlach, Valerie
; APPLICANT: Gilbert, Jennifer A.
; APPLICANT: Gunther, Erik
; APPLICANT: Gorman, Linda
; APPLICANT: Guo, Xiaojia Sasha
; APPLICANT: Ji, Weizhen
; APPLICANT: Li, Li
; APPLICANT: Lin, Xiaohong
; APPLICANT: Miller, Charles E.
; APPLICANT: Millet, Isabelle
; APPLICANT: Padigaru, Muralidhara
```

Db 1566 NQIAEIERKAEKDEMEQAKRNHLRVVDSLSQTSIDAETSRNEALRVKCKQEGDLNEM 1625
Qy 244 -VORA-----REEEKREVTSHEQVTLNDIQLQMEQHNRNSKLRONMELAE----- 292
Db 1626 EIQLSHANMAEAQKQVKS-IQSLKQTOIQLDDAVRANDDLK-ENTAIVERRNNILQA 1683
Qy 293 -----LKKLIEQVE-----LREHIDKVFHKOL-----OQQLVDAKQQAQEMLKE 334
Db 1684 ELBELRAVVQETSRKLAQELIETSRVQLLHSQNTSLNQKKQMDADLSQLQTEVEE 1743
Qy 335 A--ERHOREKDFLLKEAVESQMC--ELMKOQET--HLKQQLALYTEKFEFQNTLSK 388
Db 1744 AVQECNAEBK--AKGATDAAMAEHLKQKQDSAHRLERKKNQMEQIKDLQRLDEA 1800
Qy 389 SEV-FTTFQEMEKMTKIKLEKETMYRSWESSKALLEMAEKTVRDKE-LEGLOV 446
Db 1801 EQIALGGKKQLQKLEAVRELENE-----LE-ABQK--RNAESVKGMK 1842
Qy 447 KIORLEKLCALQOTERDNLKRVODL 472
Db 1843 SERRIKELTYQTEEDRKNL-LRLQDL 1867

RESULT 49

US-10-108-260A-4080
; Sequence 4080, Application US/10108260A
; Publication No. US20040005560A1
; GENERAL INFORMATION:
; APPLICANT: HELIX RESEARCH INSTITUTE
; TITLE OF INVENTION: No. US20040005560A1e1 full length cDNA
; FILE REFERENCE: H1-A0106
; CURRENT APPLICATION NUMBER: US/10/108,260A
; CURRENT FILING DATE: 2002-03-27
; NUMBER OF SEQ ID NOS: 5458
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 4080
; LENGTH: 512
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-108-260A-4080

Query Match 8.6%; Score 233.5; DB 15; Length 512;
Best Local Similarity 24.7%; Pred. No. 3.5e-06;
Matches 115; Conservative 72; Mismatches 176; Indels 103; Gaps 17;

Qy 75 NQGG-----PGEDGACGEPAEPDAEKSRITYVARNGEPEPTPVVYGEKPSKG 122
Db 5 NQGGESADGHIQCPKPSIIGNAGEKSLSDAKKK-----KS 42
Qy 123 DPNTBEIROSDVGDHRH-----POEKKAAGLQKE--ITLLMOTLNTLSTPEKLA 175
Db 43 NRKEDDVWASGTV--KRLHKTSGECERKTKKSLSKEDLIQLLSINEGELQAREDV 100
Qy 176 LC--KYVAELLBHRNSQKMLLO--KKOSQLVQEKDHLRGHSHKAVLARSKLSLCRE 231
Db 101 LKTEKTPVELBAHYGSAEPKVLVRDLIAQKSGIGEDYVFKPISELDELEKQKE 160
Qy 232 LOR-----HNRSLKEGVQARAEFEKKEVTSFQVTLNDIQLQMEQHNRNSKLRQ 285
Db 161 TYRMLQQLLAEKCHERTVYELNEKHHTDYNKSDPTNLLEQERERLKKLQKEKA 220
Qy 286 -----NMELERLKLQEQ-----YELREHIDKVFHKOLQQLQVLD--AKLO 326
Db 221 YQARKEKENAKRLNKLDELIVLKLSPALMULVDERQMGHEQL-----GLOSQV 276
Qy 327 QAQEMLKEASERHOREKDFILKEAV-----ESQRCBLMKQOETH-----LQ 372
Db 277 EEEKLKAITSKSKEDQKLLKLVDFEHKASRPSQHEBNNAKLANQESHNRQLKLV 336
Qy 373 LYTEKPEFQNT-----LSKSSSEVFTTFQEMEKMTKIKLEKETMYRSWESSK 429
Db 337 GLTORIEELETNKNKQAE-----ELQELRDLKIAEGCGNSSLMAEVLNRKRVLE 389

Qy 430 MA-----BEKTVRDKELEGLOVKIQLEKLCALQOTERDNLKRVODL 472
Db 390 MEGKDEBITTKTESQCRBLRKLKLODEEHHSKEFRLEVEKLOCRMSEL 435
RESULT 50
US-10-080-608A-11
; Sequence 11, Application US/10080608A
; Publication No. US20030198956A1
; GENERAL INFORMATION:
; APPLICANT: Makowski, Lee
; APPLICANT: Hyman, Paul
; APPLICANT: Williams, Mark
; TITLE OF INVENTION: STAGED ASSEMBLY OF NANOSTRUCTURES
; FILE REFERENCE: 8471-010-999
; CURRENT APPLICATION NUMBER: US/10/080,608A
; CURRENT FILING DATE: 2002-02-21
; NUMBER OF SEQ ID NOS: 180
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 11
; LENGTH: 3878
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-080-608A-11

Query Match 8.5%; Score 230.5; DB 14; Length 3878;
Best Local Similarity 18.4%; Pred. No. 6.3e-05;
Matches 119; Conservative 110; Mismatches 195; Indels 223; Gaps 23;

Qy 31 GSSQAPKPEGAQARTAGSGALRDVSELSQLIEDIISTYCVNNQ-----GGPGSDG 83
Db 161 GAQDSPTHELMSELA--GKQHEI--ELANSELSEMRVITYGTGLQQORPEAAIKORDG 217
Qy 84 -----AQGEPAEPEDAESKSRITYVARNGEPEPTPVVYGEKPSKGDPT 126
Db 218 IITQLTANLQARREKQETMREFLELTQSQKLQIQFQOLQASETLRSTHSSSTAADLLQ 277
Qy 127 ---BEIROSDVGDHRHRRPOEKKAAGLQKEITLLMOTLNTLSTPEKLAALCKK---- 179
Db 278 AKQILTHQQOLEEQHLLDYQKKEDFTMQISFLQEKIKYEMEQDKKVVNSKKEIQ 337
Qy 180 -----YAEELLEHRNSOKM-----KLQKQSQLVQEKDHLRGHSHKAVLARS 223
Db 338 EKFTIIBELNTKILIEEKKTLLEKDKLTADKLKGLQEQIVQKNQETK--NMKLELTNS 395
Qy 224 K-----LESCLRELQHRN--SLKBEQVQARAEFEKKE----- 256
Db 396 KQKBSQSSEIQLMGTVLELQKRNHKDSQPETDIVQRMEOETORKLQRLAEELDEMYG 455
Qy 257 -----VTSH-----FQVTLNDIQLQME 273
Db 456 QIVQMKQBLIRQHMAQMEEMKTRHGENENALRSYSNTVNEHQIKLANVAINELNKLQ 515
Qy 274 QINERNSKLR-----QENMELAEPLKCLTEQVELREEHIDKVPK-----HK 314
Db 516 DTNSQKELKEELGHLILEKCALQRLQEDLVEELSFSEQIQARQRTIAEBSKLMEAHK 575
Qy 315 DLQQ-QLVDAKLOQQAQEMLKEAERH-----QREKDFLLKEAVESQRM--- 356
Db 576 SLSTVEDLKAETVSASERKLELEHAEVNTYKIKLEMLKEKKNVLAORMAESQAELE 635
Qy 357 -----CELAKQOE-----THLKOQL-----ALYTEKF 378
Db 636 RLRTQLLSHEEELSKLKLDELIEHRINIEKLNGLGHIHYKQIDGLQWENKSKLETWQF 695
Qy 379 EEFQNTLSKSSVFP-----TTFQEMEKMTKIKLEKETMYRSWESSK 427
Db 696 EK-DNLITKQNLILEISKDLQOOSLVNSKSEMTLQINELQKEITELR----- 744
Qy 428 LMASEKTVRDKELEGLOVKIQLEKLCALQOTERDNLKRVODLSA 474
Db 745 -QEEKKEGTLEGEVQELQKTYELLBK---QMKENDLQEKFAQLEA 787

Search completed: June 7, 2004, 14:44:21
Job time : 135 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: June 7, 2004, 14:22:20 ; Search time 46 Seconds
(without alignments)
1108.294 Million cell updates/sec

Title: US-10-023-529-8

Perfect score: 2702

Sequence: 1 KSSPQPPAGPBGAGQERSQ.....APSTEASGTGPQPTSARA 530

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 283366 seqs, 96191526 residues

Total number of hits satisfying chosen parameters: 283366

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 150 summaries

Database :

PIR 78:**

1: Piri:**

2: Piri2:**

3: Piri3:**

4: Piri4:**

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1233	45.6	676	2 JC7222	77K muscle-derived
2	375	13.9	346	2 F87844	protein T22C1.6 [i
3	372.5	13.8	335	2 T25110	hypothetical prote
4	269	10.0	992	2 T46337	hypothetical prote
5	263.5	9.8	892	2 T50985	related to transcr
6	261.5	9.7	1407	1 S28589	trichohyalin - rab
7	259	9.6	886	2 H69378	conserved hypothet
8	256	9.6	955	2 S24348	myosin heavy chain
9	256	9.5	936	2 S39083	myosin heavy chain
10	256	9.5	1938	1 JX0178	myosin heavy chain
11	256	9.5	2677	2 A38194	desmoplakin I - hu
12	253	9.4	1206	2 T34021	protein kinase SK2
13	252.5	9.3	880	2 P75103	conserved hypothet
14	251.5	9.3	1349	1 A46691	trichohyalin - she
15	249.5	9.2	1233	2 T14157	serine/threonine p
16	249.5	9.2	1392	2 A43336	microtubule-vesicl
17	249	9.2	1392	2 A43336	microtubule-vesicl
18	249	9.2	1938	2 A52993	skeletal myosin he
19	249	9.1	1001	2 T17365	serine/threonine p
20	246	9.1	1427	2 S22695	restin - human
21	245	9.1	1377	2 I38055	myosin heavy chain
22	245	9.1	1940	2 A28320	myosin heavy chain
23	243.5	9.0	993	2 S49461	synaptosomal compl
24	242	9.0	1534	2 A56734	ribosome receptor,
25	242	9.0	3259	1 A56539	giantin - human
26	240	8.9	1931	2 A59234	slow myosin heavy
27	240	8.9	2653	1 S28261	centromere protein
28	239.5	8.9	1898	1 A45973	trichohyalin - hum
29	239	8.8	4574	2 G02520	plectin - human

30	239	8.8	4684	2 A59404	plectin [imported]
31	238.5	8.8	4687	1 A39638	plectin - rat
32	237.5	8.8	1690	2 T13030	microtubule bindin
33	237	8.8	1934	2 T48153	myosin heavy chain
34	236.5	8.8	3225	2 I52300	giantin - human
35	236	8.7	978	2 A70387	conserved hypothet
36	236	8.7	1410	1 A57013	early endosome ant
37	234.5	8.7	1231	2 T18532	serine/threonine pr
38	234	8.7	1935	1 A37102	myosin beta heavy
39	234	8.7	1935	2 A59286	myosin heavy chain
40	233.5	8.6	845	2 T48176	synaptosomal compl
41	233.5	8.6	1939	1 A46762	myosin alpha heavy
42	232.5	8.6	1432	2 B85431	trichohyalin like
43	232	8.6	875	2 A23767	myosin heavy chain
44	232	8.6	1935	1 S06006	myosin beta heavy
45	231.5	8.6	1927	2 A59236	embryonic muscle m
46	231.5	8.6	1938	1 S06005	myosin alpha heavy
47	231.5	8.6	1938	2 T49464	alpha cardiac myos
48	231	8.5	1388	2 S74245	serine/threonine-s
49	230.5	8.5	1642	2 T08880	NMDA receptor-bind
50	230.5	8.5	1939	2 T48175	myosin heavy chain
51	230.5	8.5	3187	2 JC5837	364K Golgi complex
52	229	8.5	1269	2 F84730	probable myosin he
53	229	8.5	1976	2 A59252	myosin heavy chain
54	229	8.5	2442	2 T08621	centrosome associa
55	228	8.4	1164	2 T24806	hypothetical prote
56	228	8.4	1388	2 S70633	serine/threonine-s
57	228	8.4	1938	1 A40937	myosin heavy chain
58	227.5	8.4	1744	2 F86181	PI003.10 protein -
59	226.5	8.4	1790	2 S67593	transport protein
60	226	8.4	1999	1 S21801	myosin heavy chain
61	225.5	8.3	1909	2 A45592	liver stage antige
62	225.5	8.3	1939	2 T18372	repeat organellar
63	225.5	8.3	1940	1 A24922	myosin heavy chain
64	225	8.3	746	2 T47237	myosin II heavy ch
65	224.5	8.3	911	2 S51441	hypothetical prote
66	224.5	8.3	1115	2 T41342	probable coiled-co
67	223.5	8.3	1133	2 T22976	hypothetical prote
68	223	8.3	492	2 A28616	M5 protein precurs
69	223	8.3	764	2 I51302	myosin heavy chain
70	223	8.3	1974	2 T30010	hypothetical prote
71	223.5	8.2	1940	1 S04090	myosin heavy chain
72	222.5	8.2	2057	2 S61477	myosin II heavy ch
73	222	8.2	2101	2 A42184	nuclear mitotic ap
74	221.5	8.2	946	2 S28061	SCP1 protein - rat
75	221.5	8.2	1039	2 S18199	myosin heavy chain
76	221	8.2	520	2 S35575	myosin heavy chain
77	221	8.2	1091	2 T34107	hypothetical prote
78	221	8.2	1961	1 A61231	myosin heavy chain
79	220.5	8.2	2007	1 B43402	myosin heavy chain
80	220.5	8.2	2017	1 A36014	myosin heavy chain
81	220	8.1	741	2 S39082	myosin heavy chain
82	220	8.1	924	2 S06117	myosin heavy chain
83	220	8.1	2094	2 S33124	tpi protein - huma
84	219.5	8.1	852	2 D72230	conserved hypothet
85	219.5	8.1	2253	2 T30336	nuclear/mitotic ap
86	219	8.1	1804	2 T34518	myosin heavy chain
87	219	8.1	1846	2 A59289	nestin - golden ha
88	218.5	8.1	1738	2 T14867	myr 6, unconventio
89	218.5	8.1	1964	2 A59282	interaptin - slime
90	218.5	8.1	2168	2 T30171	nonmuscle myosin I
91	218	8.1	725	2 A47168	ninein - mouse
92	218	8.1	1017	2 PC4035	cardiac morphogene
93	218	8.1	1155	2 B70356	cell-cycle-depende
94	218	8.1	1313	2 A48467	chromosome assembl
95	217.5	8.0	967	1 A35075	myosin heavy chain
96	217.5	8.0	1956	2 T16416	kinesin heavy chai
97	217.5	8.0	1972	1 A41604	hypothetical prote
98	217.5	8.0	1979	1 S03166	myosin heavy chain
99	216	8.0	555	2 C96667	myosin heavy chain
100	216	8.0	1251	2 A56677	unknown protein, 7
101	216	8.0	1940	2 A59287	neuronal cell cycl
102	215.5	8.0	1034	2 T32297	myosin heavy chain
103	215.5	8.0	1034	2 T32297	hypothetical prote

103 215.5 8.0 1354 2 S69211 serine/threonine-s
104 215.5 8.0 1959 1 A33977 myosin heavy chain
105 215 8.0 484 2 S46489 M1 protein precurs
106 215 8.0 697 2 T07111 MAR binding filame
107 215 8.0 1837 2 T41023 probable nuclear p
108 214.5 7.9 2245 2 T18278 myosin heavy chain
109 214 7.9 1300 2 I53799 CGI protein - huma
110 214 7.9 1354 2 S74264 serine/threonine-s
111 214 7.9 1356 2 S32763 kinectin 1 - huma
112 214 7.9 1957 2 T38077 hypothetical coile
113 213 7.9 1875 2 S38173 myosin-like protei
114 213 7.9 1992 2 A47297 myosin heavy chain
115 212.5 7.9 1496 2 T05634 hypothetical prote
116 212 7.8 1300 2 T18364 ro-3 protein - Neu
117 212 7.8 2954 2 T14156 kinesin-related pr
118 211.5 7.8 2116 2 A26655 myosin heavy chain
119 211 7.8 1290 2 A55094 chromosomal protei
120 210.5 7.8 863 2 S37040 paramyosin - tapse
121 210.5 7.8 1313 2 F96673 hypothetical prote
122 210.5 7.8 1597 2 S68420 citron - mouse
123 210 7.8 451 2 G70241 hypothetical prote
124 210 7.8 1827 2 T16270 hypothetical prote
125 209.5 7.8 729 2 T50989 hypothetical prote
126 209.5 7.8 1039 2 S62509 probable vesicular
127 209.5 7.8 1044 2 T50213 probable vesicular
128 209 7.7 476 2 B96667 hypothetical prote
129 209 7.7 1175 2 C35815 myosin heavy chain
130 209 7.7 1201 2 A35815 myosin heavy chain
131 209 7.7 1201 2 B35815 myosin heavy chain
132 209 7.7 1270 2 T09194 adaptor protein in
133 209 7.7 1325 2 T42722 male-enhanced anti
134 209 7.7 1702 2 T14050 protein kinase [BC
135 209 7.7 2385 2 A32491 myosin heavy chain
136 209 7.7 2411 2 B32491 myosin heavy chain
137 208.5 7.7 652 2 B59102 hypothetical prote
138 208.5 7.7 1046 2 T42734 cytoplasmic linker
139 208 7.7 853 2 T51505 hypothetical prote
140 208 7.7 1029 2 H96658 hypothetical prote
141 208 7.7 1475 2 T33318 hypothetical prote
142 208 7.7 1968 1 S05697 myosin heavy chain
143 207.5 7.7 621 2 S10450 myosin heavy chain
144 207.5 7.7 1938 1 MMRW1 myosin heavy chain
145 207 7.7 734 2 T27055 hypothetical prote
146 207 7.7 1426 2 T00337 hypothetical prote
147 207 7.7 1818 1 S73952 hypothetical prote
148 206.5 7.6 583 2 C84788 probable myosin he
149 206.5 7.6 690 2 S41009 hypothetical prote
150 206.5 7.6 705 2 E88564 protein T0565.9 [i

ALIGNMENTS

RESULT 1
JC7222 77K muscle-derived protein - chicken
C:Species: Gallus gallus (chicken)
C:Date: 09-Jun-2000 #sequence_revision 09-Jun-2000 #text_change 21-Jul-2000
R:Yajima, A.; Fukui, I.; Fujimori, K.; Kiyosue, K.; Nishimune, H.; Kasai, M.; Taguchi, T.
Biochem. Biophys. Res. Commun. 269, 564-569, 2000
A:Title: MDP77: A novel neurite-outgrowth-promoting protein predominantly expressed in c
A:Reference number: JC7222; MUID:20175243; PMID:10708594
A:Accession: JC7222
A:Molecule type: mRNA
A:Residues: 1-676 <DYE>
A:Cross-references: GB:D89599; NID:G7619883; PIDN:BA94755.1; PID:G7619884
A:Experimental source: crus muscle
C:Comment: This protein, a glycoprotein and a neurite-outgrowth-promoting protein, is im
C:Keywords: coiled coil; Glycoprotein; leucine zipper; muscle
Query Match 45.6%; Score 1233; DB 2; Length 676;
Best Local Similarity 53.9%; Pred. No. 5.8e-45;

Matches 255; Conservative 84; Mismatches 100; Indels 34; Gaps 7;
QY 2 SSPGPRGPGAGPGRSQAAPAVEAECPGSSQAPRPEGAQARTAQSGALRDVSEELSR 61
DB 15 TSPQDNQG-----QSKRAEPVPSQPLSP-TNQTSAPQEPATC-----DISEELNR 59
QY 62 QLEDIILSYCYVNDNQGSGEDGA---QGEPAEPEDAEKSTTYVARNGRPEPTPVVYGEK 117
DB 60 QLEDIILKY-----GSAASLVEKEGTTAETDKPEKEDVGSMEADACEDVNEESKD 110
QY 118 EPSKDPNTEIROSDVGDHRRPQKKAKGKKEITLLMGLTNTLSTPEEKLAALC 177
DB 111 KPAEDAS-----RAKPSASKEQK--LEKKILKGKKEITLLMGLTNTLSTPEEKLDLDF 164
QY 178 KYIAELLEHRNSQOMKLLQKQSQVQKDHRLGRHSHKAVLARSKLESICRELQRNR 237
DB 165 KYIAELLEHRNAEQKQKLYLQKRAQITKQKQDQSEHSRAILARSKLESICRELQRNRK 224
QY 238 SLKEGVORAREEERKEKKEVTSHEQVTLNDILOMQEHNHNSKLRQNMELASRLKKLI 297
DB 225 TLKEETIQARAEDEKKEKKEITNHFQGLSEIQAQIEQSSERNMKLCQNTTELAKIKSI 284
QY 298 EYELREEHIDKVFQKQDLOQLQVDAKIQQAEMLKEAEERHOREKDFLKEAVESQSMC 357
DB 285 DQYELREEHLDKIFKCHRELQKLVDAKLEQSQEWMKEAEERHOREKDFLKEAVESQSMC 344
QY 358 ELMKQOETHLKOQLALYTEKEPFTONTLSKSEVFTTFKQEMKMTKIKLEKETMYR 417
DB 345 KMLKEQETVLAQAITYLSEFEFEFKTLTKSNFVFATPKQEMKMTKIKLEKETMYR 404
QY 418 SRWSSNKALLEMAEKTVRDKLEGLQVKTQRLKLCRALQTERNDLNRKVQ 470
DB 405 SRFENCNALLDMIEEKAMRTKEVCFVLKQRLNLCRALQEESENELYRKIK 457
RESULT 2
F87844 protein T22C1.6 [imported] - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C:Date: 10-May-2001 #sequence_revision 10-May-2001 #text_change 10-May-2001
R:Anonymous, The C. elegans Sequencing Consortium.
Science 282, 2012-2018, 1998
A:Title: Genome sequence of the nematode C. elegans: a platform for investigating biolo
A:Reference number: A75000; MUID:99069613; PMID:9851916
A:Note: see websites genome.wustl.edu/gsc/C_elegans/ and www.sanger.ac.uk/projects/C.el
A:Note: published errata appeared in Science 283, 35, 1999; Science 283, 2103, 1999; an
A:Accession: F87844
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-346 <STO>
A:Cross-references: GB:chr_I; PIDN:CAA99923.1; PID:G3880029; GSPDB:GN00019; CESP:T22C1.
C:Genetics:
A:Gene: T22C1.6
A:Map position: 1
Query Match 13.9%; Score 375; DB 2; Length 346;
Best Local Similarity 30.8%; Pred. No. 2.7e-09;
Matches 105; Conservative 69; Mismatches 123; Indels 44; Gaps 9;
QY 150 KGLGKEITLLMQ-----TLNLT-STPE-EKLAALCKYAEELLEHRNSQOMKLLQKQ 201
DB 3 KMFGLKISFIFNEGDEAALLSKLGVDAEKVQKLIKLAB--SEKQNAELKIKVLDYDK 60
QY 202 SOLVQEKDHLGRHNSKAVL-----ARSKLESICRELQRHNSLKEEGVQVQARAREEKKKE 256
DB 61 VVKVLDLTKELRNQNTLLRTERAKSKLEELCRGLQKANHQTREACAKMKCLEVERGL 120
QY 257 VTSHEQVTLNDILOMQEHNHNSKLRQNMELASRLKKLIQEY-----LREE 305
DB 121 AVEQLKVTLDIKTKTWAEGRSKSLAEDNKKLSKSESEIGHQYERKMKVLDIQIKKEK 180
QY 306 HIDKVFKEKDLQQLQVDAKIQQAEMLKEAEERHOREKDFLKEAVESQSMCMLKQOET 365

Db 181 YWBYGKTKDLBIKLLTAKLESASIQVKSS-----GMEKDELAKIMLEETARVGGALKTEK 236
QY 366 HLKQQLALYTKPEEFONTLSSESSEVFTFKQEMKMTKIKKLEKETTMYPSRWESSNK 425
Db 237 ALREQVQYSAKYSILTSCLSNFAFDKFDIISRVNKKCMQVKEGLSYKKSDEANK 296
QY 426 ALL-----EMAERTVDRKLEGLQVQIKQLEKLCRALQ 459
Db 297 KVLVLTMTQVYAEKTIATSDK-----KIQMLENLCRALR 330

RESULT 3

T25110
hypothetical protein T22C1.6 - *Caenorhabditis elegans*
C:Species: *Caenorhabditis elegans*
C>Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 15-Oct-1999
R:McMurray, A.
A:Accession: T25110
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-335 <WIL>
A:Cross-references: EMBL:Z75550; PIDN:CAA99923.2; GSPDB:GN00019; CESP:T22C1.6
A:Experimental source: clone T22C1
C:Genetics:
A:Map position: 1
A:Introns: 6/1; 46/3; 66/3; 91/3; 285/2

Query Match 13.8%; Score 372.5; DB 2; Length 335;
Best Local Similarity 30.8%; Pred. No. 3.3e-09;
Matches 103; Conservative 70; Mismatches 120; Indels 41; Gaps 9;
QY 152 LGKEI--TLIMOTLNTLSTPE--EKLALCKKYAELEHRNSOKMKLLQKQSOVLQVK 208
Db 1 MGNFDEALIKSL--GYPDKQVQKLIKLAB--SEKQAEIKIKLVLDYKVKVVDL 56
QY 209 DHLRGHSKAVL-----AKSKLESICRELQRNRSLSKEGVQARBEERKEVTSHFQV 263
Db 57 TEKLERNNQILLTBEAKSKLEELCRGLQKXANQHTREACAKMKKLEVERGLAVEQLKV 116
QY 264 TLNDLOLOMEQHNRNSKLROENMELAEKLIQYE-----LREEHIDKVPK 312
Db 117 TLKDIEKTAEGRSKSDSLAENKKLSKPSIGHQYEEMKVIDQIQKKEKYWESYCK 176
QY 313 HKDLOQLVDALQQAQEMLKAEERHOREKDFLLKEAVESQRMCELMKQOETHLKOQLA 372
Db 177 TKDLETKLITAKLESASIQVKSS-----GMEKDELAKIMLEETARVGGALKTEKALREQVQ 232
QY 373 LYTEKPERFQNTLSKSSSEVFTFKQEMKMTKIKKLEKETTMYPSRWESSNKALL----- 428
Db 233 EYSAKYSILTSCLSNFAFDKFDIISRVNKKCMQVKEGLSYKKSDEANKKVLVLT 292
QY 429 ---EMAERTVDRKLEGLQVQIKQLEKLCRALQ 459
Db 293 TNQVYAEKTIATSDK-----KIQMLENLCRALR 319

RESULT 4

T46337
hypothetical protein DKFZp43402413.1 - human (fragment)
C:Species: *Homo sapiens* (man)
C>Date: 04-Feb-2000 #sequence_revision 04-Feb-2000 #text_change 04-Feb-2000
R:Koehler, K.; Beyer, A.; Mewes, H.W.; Gassenhuber, J.; Wiemann, S.
A:Accession: T46337
A:Status: preliminary
A:Molecule type: mRNA

A:Residues: 1-992 <AAA>
A:Cross-references: EMBL:AL137265
A:Experimental source: adult testis; clone DKFZp43402413
C:Genetics:
A:Note: DKFZp43402413.1

Query Match 10.0%; Score 269; DB 2; Length 992;
Best Local Similarity 21.9%; Pred. No. 0.00022;
Matches 133; Conservative 102; Mismatches 208; Indels 164; Gaps 20;

QY 7 PEAGPEGAQERPSQAAPAEVAREGP-----GSSQAPRKEPGAQARTAQSGALRDYSEELSR 61
Db 104 EKADPTG-----SRPAKASEKEAPEPTVDAGEEGSRREERAKPKKASALEGSSDSASQ 158
QY 62 QLE-----DILSTV-----CYDNNQGG 78
Db 159 ELEISEHMKPEQLSDSIASDPKSPHGLDFGFRSRISEHLDDVDVLSPVLGACACQAQAPL 218
QY 79 PGEDGAQGEPAEPZ--DAKSRITYVARGEP-----BPTPVYVGEKPSKGPNTTEEI 129
Db 219 GIEDKDSQSSQDELSQSKGLERLSPLPHEERAQSPRSLATEEPPQSGEGQPEW 278
QY 130 QSDDEVGD-----RDHRRPQEKKKA 149
Db 279 KEARELGDSASLSLSQLSLOEQAPSPPAACEKQKQHSQAESLGPQOEAEDEPEKVA 338
QY 150 KGLGKEITLLMOTLNTLSTPE-----KLAALCKKYAELEHRNSOKMKLLQKQSOVLV 205
Db 339 VSPTPPVSPVRSSTEPVAPPQLSEALKAMEEAVQVLE-----QDQRHLLESKQERTMQ 393
QY 206 QSKDHLRGHSKAVLA--RSKLESICRELQRNRSLSKEGVQARBEERK-----RKEVT 258
Db 394 QIREKLCQEEETIRLHQKEQLSSLSRERLQKAIHEEA--RMREESQRLSWLRQVQ 452
QY 259 SHFQVTLNDILOME-----QHNRNSKLQENMELAEKLIQYEVELREH 306
Db 453 SSTQADEQIRAEQBSLQKLEBESQKAEASLEQKNQMLEQLKEEIEASEKSEQA 512
QY 307 IDKVPKHLQLOQ---QLVDALQQAQEMLKE-----AEERHOREKDFLLKEAVE 352
Db 513 ALNAAKALQQLREQLEGERKEAVATLEKESHALEKLCSSLEAKHREVVSSLOKQIQE 572
QY 353 SQRMCELMKQ---QETHLKQALALYTEKFEFFQNTL--SKSSEVFTFKQEMKMTKIK 406
Db 573 AQKKEAQLQCLQVHERVHQSYHVAGYHEHLSLLREKQVGEGBEHERLDKM----- 628
QY 407 KKLKETTMYPSRWESSNKALLEMAEKTVDRKLEGLQVQIKQLEKLC--RALQTERNDL 465
Db 629 -KEHQVQVMAKAREQYB-----AEERKQAEILLGHLTGELERLQRAHARELETVRQEQ 680
QY 466 NKRVDQL 472
Db 691 HKRLIEDL 687

RESULT 5
T50985
related to transcription factor TMF [imported] - *Neurospora crassa*
N:Alternate names: protein B7F18.40
C:Species: *Neurospora crassa*
C>Date: 21-Jul-2000 #sequence_revision 21-Jul-2000 #text_change 21-Jul-2000
R:Schulte, U.; Aign, V.; Hoheisel, J.; Brandt, P.; Partmann, B.; Holland, R.; Nyakatura, submitted to the Protein Sequence Database, July 2000
A:Accession: T50985
A:Reference number: 225286
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-892 <SCH>
A:Cross-references: EMBL:AL389891; GSPDB:GN00116; NCSP:B7F18.40
A:Experimental source: BAC clone B7F18; strain OR74A
C:Genetics:
A:Gene: NCSP:B7F18.40

Db 215 -----EVRNLESLKELBEHKSRL-----ESLRKQSSVLCVGRGLBEKLELE-- 259
Qy 236 NRSLEKGVQARBEERKEV-----TSHFQVTLND----- 267
Db 260 -KQLKE-VERIEBLEKAKVEKELPKAARYSILEKLLSEINQALADVEKREGDLTREA 317
Qy 268 --LOLMEQHNRNRSKLOENMELAEKLLILOYE-----IR 303
Db 318 AGIOAQKKAEDNSKL-EETKRIEELEREFKSHRLLETLPKPMORMQGIKAKLE 376
Qy 304 EEHf--DKVFHKDLOQLVDKALQOQEMLEKABERHOR---EKDFL-----LKEAVE 352
Db 377 EKNLTPDKVEKMYDL-----LSKAKBEKEITEKXKLAKKSSLKTRGAQLKKA 428
Qy 353 -----SQRCMLMKQO--ETHLKQALALYTEKFEFQNTLSKSSSVFTTFQKEMKMTYKI 406
Db 429 ELKSAERTCPVCGELDEEHRKNIMAYETREMKRIABELAKADIEKKLERLSKVE--- 485
Qy 407 KLEKETMYRSRW-----SSNKALLEMABEKTVRDKELEGL--QVKI-- 448
Db 486 KALEKQETVLKYRWVDELKALENELSHDAKLSAESSEYRKVKER-LDGLRQCKILL 544
Qy 449 -----QRLSKLCRALQTERNDLNKRVOD 471
Db 545 SSASRIKELXSSLEIEALKVNESRGELHRKIRE 580

RESULT 8
S24348
myosin heavy chain, embryonic and adult skeletal muscle (clone Cemb2) - chicken (fragment)
C:Species: Gallus gallus (chicken)
C>Date: 03-Feb-1994 #sequence_revision 06-Sep-1996 #text_change 13-Feb-1998
C:Accession: S24348
R:Moore, L.A.; Arrizubieta, M.J.; Tidyman, W.E.; Herman, L.A.; Bandman, E.
J. Mol. Biol. 225, 1143-1151, 1992
A:Title: Analysis of the chicken fast myosin heavy chain family. Localization of isoform
A:Reference number: S24348; MUID:92309413; PMID:1377278
A:Accession: S24348
A:Molecule type: mRNA
A:Status: nucleic acid sequence not shown
A:Residues: 1-955 <MOO>
A:Cross-references: EMBL:M74085
A:Superfamily: myosin heavy chain; myosin motor domain homology
C:Keywords: ATP; coiled coil; muscle contraction; skeletal muscle; thick filament

Query Match 9.6%; Score 259; DB 2; Length 955;
Best Local Similarity 22.6%; Pred. No. 0.00055;
Matches 118; Conservative 82; Mismatches 181; Indels 142; Gaps 16;
Qy 33 SOAPRKEGQAQTAQSGALRDVSELSRQLEDILSTYCVNNQGGPGEQAQEPAPPE 92
Db 107 SQIQSKIEDQALGMQ---LQKKIKELQARIEL-----ESEEIERTSRA 149
Qy 93 DAERSRTYVARNGEPEPTPVVYGEKPSKDPNTEIRQSDVEGDRDHRFPQKCAKGL 152
Db 150 KAEKHRADLSLEEE-----ISELEAGGATATQI-----DMNKEAEFQWRD 196
Qy 153 GKEITLMTLNTUSTPEEKALCKKY-----AELEBHRNSQKQMLKQKQSQLVQEK 208
Db 197 LEEATL-----QHEATAAALRKHADSTAEGLQDNLQVRVKQKLEKESKELMCI 247
Qy 209 DHLRGKSHKAVLARSKLESICREL-----QRNRSLEKEGVQVOR-- 248
Db 248 DDLASNMESVSKANLEKMCRTLEDQLSKISEBEHQRMINDLSTQRLQTESGEYS 307
Qy 249 -----BEEKKEVTSH-FQVTLNDILOMEQHN 276
Db 308 QVBEKDALISQSRGKAFTQQIEELKRLHEEIEKAKNALAHALQSAHDCDILLREQYE 367
Qy 277 ER-----NSKL-----QENMELAEKLLILOYELREHIDKVP 311
Db 368 EQEAKGELQALSKANSEVAQWTKYTETAIQTEELBEAKKLAQLQDASEHV---- 423

Qy 312 KHKDLOQLVDKALQOQEMLEKABERHORQEKDFLLKEAVESQRMCELMKQETHKQQL 371
Db 424 -----EAVNAKAS-----LEKTKQRLQNEVEDLMIDVERANAACARLDKKQNFDKIL 472
Qy 372 ALYTEKFEFQNTLSKSSSVFTTFQKEMKMTYKI KLEKETMYRSRVSSNKALLEMA 431
Db 473 AEKWKQYETQAELEASQKESRSLSTELFM-----KNAYEESLHLETLKRNKNLQOEI 528
Qy 432 SEKTVDKBEGLQVQKIQBLEKLCRALQTERNDLNKRVQDLSA 474
Db 529 SOLT--EQIAEGKA-IHELEKVKQIEQEKSEIQALAEAEA 568

RESULT 9
S39083
myosin heavy chain, neonatal [similarity] - chicken (fragment)
C:Species: Gallus gallus (chicken)
C>Date: 03-Feb-1994 #sequence_revision 03-Feb-1994 #text_change 20-Jun-2000
C:Accession: S39083; S24350; A26821
R:Moore, L.A.; Arrizubieta, M.J.; Tidyman, W.E.; Herman, L.A.; Bandman, E.
submitted to the EMBL Data Library, August 1991
A:Description: Analysis of the chicken fast myosin heavy chain family: Localization of
A:Reference number: S39081
A:Accession: S39083
A:Molecule type: mRNA
A:Residues: 1-936 <MOO1>
A:Cross-references: EMBL:M74087
R:Moore, L.A.; Arrizubieta, M.J.; Tidyman, W.E.; Herman, L.A.; Bandman, E.
J. Mol. Biol. 225, 1143-1151, 1992
A:Title: Analysis of the chicken fast myosin heavy chain family. Localization of isoform
A:Reference number: S24348; MUID:92309413; PMID:1377278
A:Accession: S24350
A:Molecule type: mRNA
A:Residues: 3-466, 'Q', 468-641, 'R', 643-936 <MOO2>
A:Cross-references: EMBL:M74087
R:Moraricity, D.M.; Barringer, K.J.; Dodgson, J.B.; Richter, H.E.; Young, R.B.
DNA 6, 91-99, 1987
A:Title: Genomic clones encoding chicken myosin heavy-chain genes.
A:Reference number: A26821; MUID:87217964; PMID:3034534
A:Accession: A26821
A:Molecule type: DNA
A:Residues: 'F', 856-936 <MOR>
A:Cross-references: GB:M16557; NID:9212371; PIDN:AAA48970.1; PID:9212372
C:Genetics: 886/3
A:Introns: 886/3
C:Superfamily: myosin heavy chain; myosin motor domain homology
C:Keywords: actin binding; ATP; coiled coil; muscle; skeletal muscle

Query Match 9.5%; Score 256; DB 2; Length 936;
Best Local Similarity 22.0%; Pred. No. 0.00072;
Matches 122; Conservative 80; Mismatches 180; Indels 172; Gaps 18;
Qy 19 SQAAPAVEAGFGSSQAPRPEGQAQTAQSGALRDVSELSRQLEDILSTYCVNNQGG 78
Db 89 SQIQSKIEDQALGMQKQKIKELQART-----EELEEEIE- 124
Qy 79 PUEGQAGPEAPEDAKSRITVARNGEPEPTPVVYGEKPSKDPNTEIRQSDVEGDR 138
Db 125 -----AERTSRKAEKHRADLSLEEB-----ISELEAGGATAAQI-----DM 164
Qy 139 DHRPQEKKAKGLGKIEITLMTLNTLSTPEEKALCKKY-----AELEBHRNSQKQ 194
Db 165 NKKREAEFQWRDLEBATL-----QHEATAAALRKHADSTAEGLQDNLQVRVK 215
Qy 195 KLLQKQSQLVQEKQHLRGEHSHKAVLARSKLESICREL-----QRNRSLEKEG 243
Db 216 QKLEKESKELMCI DDLASNMESVSKASLEKTKCTRALEDQMSIEKTEKEHQRMINDVN 275
Qy 244 VQAR-----BEEKKEVTSH-FQ 262
Db 276 AQARLQTESGEYSQVBEKDALISQSRGKAFTQQIEELKRLHEEIEKAKNALAHGLQ 335
Qy 263 VTINDILOMEQHN-----NSKL-----QENMELAEKLLI 297

Db 336 SARHDCDLLREQVEEQEAKGELQALSKANSEVAQWRKTYETDAIQRTSELEAKKLA 395
QY 298 EOVELREEHIDKVPFKHDLQOQAVDAKLOQAQEMLEKABERHOREKDFLLKEAVESQEMC 357
Db 396 QRLQDABEHV-----EAVNSKAS-----LETKORLQNVEDLMIDVERANSAC 440
QY 358 ELMKQOQTHLKQALALYTKFPEPONTL-----SKSEVF---TTFKQ---EMEKMT 403
Db 441 AALDKQKQNFDKILSEFKQYETQABLEASQKESRLSTELFKMKNAYEESLDHLETLK 500
QY 404 KIKLKLEKETMYRSWESSNKALLEMAEKTVDKLEGLQVYKIQLE-----KLC 455
Db 501 RENKLNQQLSULTEQIABGKKAHELEKVKQIQEQKSLQASLEAREASLEHBECKIL 560
QY 456 RALQTERNDLNRV 469
Db 561 R-LQLELNQVNSEI 573

RESULT 10
JX0178
myosin heavy chain, fast skeletal muscle, adult [validated] - chicken
N:Contains: myosin ATPase [EC 3.6.4.1]
C:Species: Gallus gallus (chicken)
C:Date: 31-Dec-1993 #sequence revision 31-Dec-1993 #text change 19-Apr-2002
C:Accession: PX0050; PX0051; PX0052; JX0178; A26365; S02082; PW0009; S39081; S24351; S05
R:Hayashida, M.; Maita, T.; Matsuda, G.
J. Biochem. 110, 54-59, 1991
A:Title: The primary structure of skeletal muscle myosin heavy chain: I. Sequence of the
A:Reference number: PX0050; MUID:92041767; PMID:1939027
A:Accession: PX0050
A:Molecule type: protein
A:Residues: 1-205 <RAY>
R:Komine, Y.; Maita, T.; Matsuda, G.
J. Biochem. 110, 60-67, 1991
A:Title: The primary structure of skeletal muscle myosin heavy chain: II. Sequence of th
A:Reference number: PX0051; MUID:92041768; PMID:1939028
A:Accession: PX0051
A:Molecule type: protein
A:Residues: 206-636 <KOW>
R:Maita, T.; Miyaniishi, T.; Matsuzono, K.; Tanioka, Y.; Matsuda, G.
J. Biochem. 110, 68-74, 1991
A:Title: The primary structure of skeletal muscle myosin heavy chain: III. Sequence of b
A:Reference number: PX0052; MUID:92041769; PMID:1939029
A:Accession: PX0052
A:Molecule type: protein
A:Residues: 201-213;632-837 <MAI>
R:Maita, T.; Yajima, E.; Negata, S.; Miyaniishi, T.; Nakayama, S.; Matsuda, G.
J. Biochem. 110, 75-87, 1991
A:Title: The primary structure of skeletal muscle myosin heavy chain: IV. Sequence of th
A:Reference number: JX0178; MUID:92041770; PMID:1939030
A:Accession: JX0178
A:Molecule type: protein
A:Residues: 833-1938 <MA2>
R:Maita, T.; Hayashida, M.; Tanioka, Y.; Komine, Y.; Matsuda, G.
Proc. Natl. Acad. Sci. U.S.A. 84, 416-420, 1987
A:Title: The primary structure of the myosin head.
A:Reference number: A26365; MUID:87092420; PMID:3467365
A:Accession: A26365
A:Molecule type: protein
A:Residues: 1-129, 'X', 131-139, 141-550, 'X', 552-754, 'X', 756-784, 'QL', 787-804, 806-810 <MA3>
R:Watanabe, B.
Biol. Chem. Hoppe-Seyler 370, 55-61, 1989
A:Title: Amino-acid sequence of the hinge region in chicken myosin subfragment-2.
A:Reference number: S02082; MUID:89228549; PMID:2713098
A:Accession: S02082
A:Molecule type: protein
A:Residues: 1144-1270 <WAT>
R:Yajima, E.
Nagasaki Igakkai Zasshi 65, 409-430, 1990
A:Title: Study on tail region of skeletal muscle myosin; primary structure and protease
A:Reference number: PW0009

A:Accession: PW0009
A:Molecule type: protein
A:Residues: 1304-1938 <VAY>
R:Moore, L.A.; Arrizubieta, M.J.; Tidyman, W.E.; Herman, L.A.; Bandman, E.
submitted to the EMBL Data Library, August 1991
A:Description: Analysis of the chicken fast myosin heavy chain family: Localization of
A:Reference number: S39081
A:Accession: S39081
A:Molecule type: mRNA
A:Residues: 1081-1203, 'DV', 1206-1342, 'E', 1344-1544, 'S', 1546-1795, 'HV', 1798-1829, 'S', 183
A:Cross-references: EMBL:M74084
R:Moore, L.A.; Arrizubieta, M.J.; Tidyman, W.E.; Herman, L.A.; Bandman, E.
J. Mol. Biol. 225, 1143-1151, 1992
A:Title: Analysis of the chicken fast myosin heavy chain family. Localization of isofo
A:Reference number: S24349; MUID:92309413; PMID:1377278
A:Accession: S24351
A:Molecule type: mRNA
A:Residues: 1082-1182, 'T', 1184-1203, 'DV', 1206-1342, 'E', 1344-1544, 'S', 1546-1780, 'M', 1782
A:Cross-references: EMBL:M74084
R:Watanabe, B.
Biol. Chem. Hoppe-Seyler 370, 1027-1034, 1989
A:Title: Complete amino-acid sequence of subfragment-2 in adult chicken skeletal muscle
A:Reference number: S05515; MUID:90121764; PMID:2610940
A:Accession: S05515
A:Molecule type: protein
A:Residues: 842-906, 'Q', 908-1270 <WA3>
R:Watanabe, B.
Biol. Chem. Hoppe-Seyler 370, 549-558, 1989
A:Title: Amino-acid sequence of the short subfragment-2 in adult chicken skeletal muscul
A:Reference number: S04501; MUID:89374803; PMID:2775482
A:Accession: S04501
A:Molecule type: protein
A:Residues: 852-906, 'Q', 908-1108 <WA2>
R:Matsuda, G.; Maita, T.; Miyaniishi, T.; Hayashida, M.
J. Protein Chem. 6, 33-46, 1987
A:Title: Structure and function of muscle myosin.
A:Reference number: A60877
A:Accession: A60877
A:Molecule type: protein
A:Residues: 1-139, 141-205 <MA1>
R:Gulick, J.; Kropp, K.; Robbins, J.
J. Biol. Chem. 260, 14513-14520, 1985
A:Title: The structure of two fast-white myosin heavy chain promoters. A comparative st
A:Reference number: A32507; MUID:8603956; PMID:2997212
A:Accession: A24124
A:Molecule type: DNA
A:Residues: 'M', 1-168 <GUL>
A:Cross-references: GB:M13512; GB:M12083; GB:M13510; NID:G212363; PIDN:AAA48966.1; PID:
R:Kropp, K.; Gulick, J.; Robbins, J.
J. Biol. Chem. 261, 6613-6618, 1986
A:Title: A canonical sequence organization at the 5'-end of the myosin heavy chain gene
A:Reference number: A32587; MUID:86196091; PMID:3009465
A:Accession: C25217
A:Molecule type: DNA
A:Residues: 'M', 1-56, 'T', 58-76, 'I', 78-168 <KRO>
A:Cross-references: GB:M13515; GB:M13511; NID:G212373; PIDN:AAA48971.1; PID:G555468
C:Comment: This is a fragment of the globular head.
C:Superfamily: myosin heavy chain; myosin motor domain homology
C:Keywords: acetylated amino end; actin binding; ATP; coiled coil; hydrolyase; methylate
F:1-1938/Product: myosin heavy chain #status experimental <MAT>
F:89-768/Domain: myosin motor domain homology <MYOT>
F:179-186/Region: nucleotide-binding motif A (P-loop)
F:550-587/Region: actin binding #status predicted
F:657-679/Region: actin binding #status predicted
F:841-1938/Domain: coiled coil <COI>
F:841-1289/Region: S2
F:852-1108/Domain: short subfragment 2 <SUB2>
F:1290-1938/Region: light meromyosin
F:1/Modified site: acetylated amino end (Ala) #status experimental
F:35/Modified site: N6-methyllysine (Lys) #status experimental
F:130,551/Modified site: N6,N6-trimethyllysine (Lys) #status experimental
F:185/Binding site: ATP (Lys) #status predicted
F:697,707/Active site: Cys #status predicted

F:755/Modified site: 3'-methylinhistidine (His) #status experimental

Query Match 9.5%; Score 256; DB 1; Length 1938;
Best Local Similarity 22.4%; Pred. No. 0.0015;
Matches 121; Conservative 83; Mismatches 178; Indels 159; Gaps 18;

QY 33 SOAPKPEGAQARTAGSALRVSVLSRQLDILSTCYVDNNGGPGEDGAGQGPAPPE 92
DB 1090 SOIQKIEDEQALGQV---LQKKIHELQRIEEL-----EEIEAERTSRA 1132
QY 93 DAEKRTYVARNGEPEPTPVVYGEKPSKGDPTNTEIRQSDVGVDRHRRPOEKKAKGL 152
DB 1133 KAEKRRADLSRELE-----ISERLEAGGATAAQI-----EMTKREAFQKMRD 1179
QY 153 GKEITILLMTLNTLSTPEKLAALCKY---AELESHENSOKMKLQKQSOLOVSK 208
DB 1180 LEEATL-----QHEATAALRRKHADSTAEIGQIDNLRVQKLEKSELMKEI 1230
QY 209 DILRGHSHKAVLARSKLESICREL-----QRNRSKESGVQVQAR----- 248
DB 1231 DDLASNESVSQAKANLEKQWCTLEDQLSEIKTKEEQNRMINDLNTQARLQETGEYS 1290
QY 249 -----BBEKKKEVTSH-FQVTLNDILOMEQH 276
DB 1291 QRAEBKDALISQSRGKQFTQIIBELKXHLSEETKAKVALAHALQSAHDCDILREQYE 1350
QY 277 ER-----NSKLR-----QENNELAERLKLIEQVELFEHIDKVF 311
DB 1351 EQEAKGELQALRSANSEVQWRTKYETDAIQRTTELEAKKLAQLQDAEHV----- 1406
QY 312 KHKDLOQQLVDAKLAQOQENKAEERHOREKDFLKEAVESQRMCKMQQETHLKOOL 371
DB 1407 -----EAVNAKAS-----LETKQRLQNEVEDLMDVVERSNAAACALDKKQKFPKIL 1455
QY 372 ALYTEKFEFPONTL-----SKSEVF---TTPKQ---EMEKMTKKIKLEKETTMYR 417
DB 1456 AKWKQIBETQTELEASQKESLSSTELFPWKNVAYEESDLHETLKREKNKLAQELADLT 1515
QY 418 SRWESSNKALLEMAEKTVDKLEGLQVKIQRL-----KLCRALQTERNDLNKRV 469
DB 1516 EQIAEGGKVAHELVKVKHVEQKESBLQALBEAASLSHERGKILR-LQLELNQTKSEI 1574

RESULT 11
A38194
desmoplakin I - human
N:Contains: desmoplakin II
C:Species: Homo sapiens (man)
C>Date: 04-Mar-1993 #sequence revision 03-May-1996 #text_change 21-Jul-2000
C:Accession: A38194; A35536; B35536
R:Virata, M.L.; Wagner, R.M.; Parry, D.A.; Green, K.J.
Proc. Natl. Acad. Sci. U.S.A. 89, 544-548, 1992
A:Title: Molecular structure of the human desmoplakin I and II amino terminus.
A:Reference number: A38194; MUID:92115697; PMID:1731325
A:Accession: A38194
A:Molecule type: mRNA
A:Residues: 1-974 <VR>
A>Note: sequence extracted from NCBI backbone (NCBI:75984, NCBI:75986)
R:Green, K.J.; Parry, D.A.D.; Steinert, P.M.; Virata, M.L.A.; Wagner, R.M.; Angst, B.D.; J. Biol. Chem. 265, 2603-2612, 1990
A:Title: Structure of the human desmoplakins. Implications for function in the desmosome
A:Reference number: A35536; MUID:90153880; PMID:1689290
A:Accession: A35536
A:Molecule type: mRNA
A:Residues: 1-927-2677 <GR>
A:CROSS-references: GB:J05211; NID:g181607; PIDN:AAA35766.1; PID:g181608
A:Accession: B35536
A:Molecule type: mRNA
A:Residues: 1-927-1000, 1600-2677 <GR2>
A:CROSS-references: GB:J05211
C:Comment: Desmoplakins I and II, products of single gene, are major proteins of the int
C:Genetics:
A:Gene: GDB:DSF

A:Cross-references: GDB:126564; OMIM:125647
A:Map position: 6pter-6p21
C:Keywords: alternative splicing; cytoskeleton
F:1-2677/Product: desmoplakin I #status predicted <MAT1>
F:1-1000,1600-2677/Product: desmoplakin II #status predicted <MAT2>

Query Match 9.5%; Score 256; DB 2; Length 2677;
Best Local Similarity 23.2%; Pred. No. 0.0021;
Matches 133; Conservative 107; Mismatches 195; Indels 138; Gaps 25;

QY 26 EABPGSGSAPRPK-----BGAQARTAS-----GALRVSHLSRQLEDIL 67
DB 1238 QOKATGSEVSQKQOQLEVLQVQMTESVRYKQSLDDAAKTIQOKNKIEIR-LKOLI 1296
QY 68 STYCVNNGGPGEDGAGPEDEDAEKST-VVANGPEPEPTVV-----YGEKPSKG 122
DB 1297 DKEINR-----KLEENARLQVQDLOKANSSATETINKLAVQBOELREL 1344
QY 123 DPNTETIRQSDVGVDRHRRPOEKKAKGLQKEITLLMTLNTLSTPEKLAALCKYAE 182
DB 1345 RIDYERVSVQRTVADQDITRFQNSLKELOLQK--KVEEELNRLKRTASEDSCKRKLLE 1402
QY 183 LLESHENSOKMKL-----LQKKQSO--LVQEKD-----HLR-GEHSHKAVLAR 222
DB 1403 ELEGRRSLKEQAIKITNLTOOLEQASIVKRSSEDDLRQDRDVLGHLREKQTOEELRR 1462
QY 223 --SKLESICRELQHRNLSKE-----EGVQARBEERKEKKEVTSFQVTLNDILOMEQH 275
DB 1463 LSSEVEALRRQLQOESVQKQHLRNHFPQAIKDKSR-----SLNESKTEIERL 1512
QY 276 NERNKLAQENKELAEKLI-IEQYELR-----BEHIDK-----VFKHKLQOQ 318
DB 1513 QSLTENLTKEHLMLEELNRLYDDLRGRRSADSDKATILELSQLOIINNRTLEL 1572
QY 319 QLVPAKLAQOQAEMLKEAERHOREKDFLKEAVESQRMCK-ELMKQOETHLQKQALYTEK 377
DB 1573 QGLINDLORENLARQETEFQKQALASNRIOESKNOCTQVQVQERESLLVKKIVLEQDK 1632
QY 378 -----PEFPQNTLSKSEVFPTPOEMKMTKKIKK-LEKETTMYRSWESSNKALLEMA 431
DB 1633 ARLOQLEDELNRAKSTLEAETRVKORLECEKQOIQNDLNQWKTQYSRKEAIRK--IESE 1690
QY 432 BEKTVRDK-----ELEGLQVKIORLEKLCR-----ALQTERNDLNKRVQDLISA 474
DB 1691 REKSEKNSLSRSEIERLQAEIKRIEERCKRKLSDSTRETQSQLETSRYQREIDKL-- 1748
QY 475 GQGSGLTDSGPRPEPGCAQAPSPRVTEAPC 507
DB 1749 -----RQRPYG-----SHRETQTEC 1763

RESULT 12
T34021
protein kinase SK2 - rat
C:Species: Rattus norvegicus (Norway rat)
C>Date: 29-Oct-1999 #sequence_revision 29-Oct-1999 #text_change 29-Oct-1999
C:Accession: T34021
R:Fukami, Y.; Yamamoto, H.; Ichihara, T.; Mori, K.; Gomi, T.; Sato, K.
Submitted to the EMBL Data Library, April 1997
A:Description: SK2, a putative rat homologue of yeast protein kinase NRK1.
A:Reference number: Z21463
A:Accession: T34021
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 1-1206 <FK>
A:CROSS-references: EMBL:AB003357; NID:d1106513; PID:d1020890; PIDN:BAA20077.1
A:Experimental source: strain SD
C:Genetics:
A:Gene: SK2

Query Match 9.4%; Score 253; DB 2; Length 1206;
Best Local Similarity 22.2%; Pred. No. 0.0013;
Matches 133; Conservative 92; Mismatches 199; Indels 176; Gaps 25;

QY 12 EGQERPSQAAPVAEAGPSSQA-----PR 37
Db 682 ENAQELPVKAEP-----QAPAAQASEPPVLPISINHSNTENKMGALPKETILPP 737
QY 38 KPEGAQARTASG---ALRDSVSELSQLEIDILSTYCDVNNQGGPGEAGQEPASPEDA 94
Db 738 EPENGKNDTSGTGVSNSSDLNLSISFSL-----KTKDSGSVSIQETRRQKTL 791
QY 95 EKRTYVARNGEPEPTPVVYGEKPSKGDNTBEIROSDVGDORDHRRPOEKKKAGLGLK 154
Db 792 KTKRFTVDGVEVSVT-----SKIVTSDSKTEBLR-----FLRQELSELRLLOK 838
QY 155 EITLLMOTLN-TLSTPEKLAALCKYAELEHNSQOMKLLQKQSOLOVQKDHRLG 213
Db 839 EEQAQOOLNKLQOQEQIFRFEQ--EMLSKKQYDQIENLEKQKXTIB-----889
QY 214 EHSKAVLARSKLSCLRELQHNHNSLKEGVORARBEERKKEVTSHPVOTLNDIQLOME 273
Db 890 -----RLE-----QBHTNRLRDE-AKRIKGEQKE-----LSKFNWMLRN-----923
QY 274 QHNERNKSLKQNNMELAKLLEQ-----YELREEH 306
Db 924 RXGEEQFVQOQOELQALKKIIOQKQAEALNTERECLNNKQOLLARAEAAWLEERH 983
QY 307 IDKVPKHDLQOQQLVDKLAQQAQEMKEAE---ERHOREKDFLLKEAV--ESQSMCELMK 361
Db 984 LQE--KHQLLAQQLKQVFIQRHQLLKHEKETEOMQRYNQRLTEELKNRQTOBRARLPK 1041
QY 362 QOETHLQOALY-----TEKFEFQNTLSKSEVFTTFKQEMKTKKI 406
Db 1042 IORSEAKTRAMFKKSLRINSTATPDQDREKIKQFAAQEEK-----ROKNRMAQH- 1092
QY 407 KKLEKTTMYRSWESSNKALLEMAEEK-----TVDRKELE---GLQVKIQRLEKLC 455
Db 1093 QKHESQMRDLQOCEANVRELHOLQNEKCHLVEHETQKLKELSEHSQELKEWR-EKLR 1151
QY 456 RALQTERNDLNKRVODLS-----AGGQSLSDSGPERRPQGAQSSPVTEARCPYG 510
Db 1152 PRKKTLEEFARKLQEQEVFFMTGSECLNPSAQSR-----GCLQTSHPSSTEAPAWAG 1206

RESULT 13
F75103
conserved hypothetical protein PAB0812 - Pyrococcus abyssi (strain Orsay)
C:Species: Pyrococcus abyssi
C:Date: 20-Aug-1999 #sequence_revision 20-Aug-1999 #text_change 02-Feb-2001
C:Accession: F75103
R:anonymous, Genoscope
submitted to the EMBL Data Library, July 1999
A:Description: Pyrococcus abyssi genome sequence: insights into archaeal chromosome str
A:Reference number: A75001
A:Accession: F75103
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-880 <KAW>
A:Cross-references: GB:AJ248286; GB:AL096836; NID:9545836; PIDN:CAB50131.1; PID:9545864
A:Experimental source: strain Orsay
C:Genetics:
C:Superfamily: Archaeoglobus fulgidus conserved hypothetical protein AF1032

Query Match 9.3%; Score 252.5; DB 2; Length 880;
Best Local Similarity 23.2%; Pred. No. 0.00095;
Matches 123; Conservative 98; Mismatches 183; Indels 127; Gaps 22;

QY 12 EGQERPSQAAPVAEAGPSSQ-----APRKPEGQAQRTAQGALRDVSELSRQLE 64
Db 259 KGLEEKIVQIERSIEBKAKISELEIVNDIPKLOEKEVYKRLKG-FRDEYESKLRLLE 317
QY 65 DILSTYCDVNNQGGPGEAGQEPAPEDAESKRTYVARNGEPEPTPVVYGEKPSKGD 124
Db 318 KELSWM-----ESELNAIEVYKGR-----KKKE 342

QY 125 NTEIROQS-DEVGDR-DHRRP--QEKKAQKGLKEITLIMQTLNLTSTPE--EKLAALCK 178
Db 343 RAEEIREKLSIEKLEELHLPVVELEDAKQVQKIERKARLKGSLSPGEVIEKLESLEK 402
QY 179 KYAELEHNSQOMKLLQKQSOLOVQKDHRLHREHSAKAVLARSKLSLC-----REL-Q 233
Db 403 ERTEI-----EBAKEITTRIGOMEQEK-----ERMKAIBELKAKGKCPVQRELTE 451
QY 234 RHNRSLK-----REGVORAREEERKEVTSHPVOTLNDIQLQ---MEQHNRNS 280
Db 452 EHKELMERYTLEIKIBELARTTEERKLVNRLKLEIKLREPSVMDIAEQIKELSES 511
QY 281 KLROENMELAE-----RLKGLIEQYELREHIDKVPK 312
Db 512 KLKGFNLELEQKEREFEGLNEEFNKLKGEILLGLERDLKRIKALEGRRKLEEKVKRKAKE 571
QY 313 H-KDLOOQLVDKLAQQAQEM---LKEAERHOREKDFLLKEAVESQRMCELMKQOETHLK 368
Db 572 ELENHRLQELGFSVEELNRIQELSEFHDYKVEAKKSES-ELRELKNNKLEKKEKTELD 630
QY 369 QQLALYTERKFEFQNTLSKSEVFTTF-KOENKEMTKKI-KLEKETMYTRSWESSNKA- 426
Db 631 QAFENLADVENIEEKEAKLDESKEFNEEYERKERVKLEREVSSITABLEBLKSV 690
QY 427 -----LLEMAERKTVDRKEGLQVKIORLEKLCBALQTERNDLNKRVOD 471
Db 691 EQIKATIRKLKSEKEREK---AKLEIKLEKALKSVKVE---DLRKKIND 733

RESULT 14
A40691
trichohyalin - sheep
C:Species: Ovis orientalis aries, Ovis ammon aries (domestic sheep)
C:Date: 21-Sep-1993 #sequence_revision 01-Mar-1996 #text_change 22-Jun-1999
C:Accession: A40691; A34209; S32633
R:Pietz, M.J.; McLaughlan, C.J.; Campbell, M.T.; Rogers, G.E.
J. Cell Biol. 121, 855-865, 1993
A:Title: Analysis of the sheep trichohyalin gene: potential structural and calcium-bind
A:Reference number: A40691; MUID:93260018; PMID:7684041
A:Accession: A40691
A:Molecule type: DNA
A:Residues: 1-1549 <PIE>
A:Cross-references: EMBL:Z18361; NID:g295940; PIDN:CAA79165.1; PID:g295941
A:Note: sequence extracted from NCBI backbone (NCBIP:132511)
R:Pietz, M.J.; Presland, R.B.; Rogers, G.E.
J. Cell Biol. 110, 427-436, 1990
A:Title: The cDNA-deduced amino acid sequence for trichohyalin, a differentiation marke
A:Reference number: A34209; MUID:96130632; PMID:2298812
A:Accession: A34209
A:Molecule type: mRNA
A:Residues: 1016-1151, 1205-1257, 1281-1398, 'G', 1400-1549 <PI2>
A:Cross-references: GB:X51695; NID:g1827; PIDN:CAA35992.1; PID:g1828
C:Comment: trichohyalin is a protein of the medulla of the hair and of the inner root s
Covalent modifications to this protein include conversion of arginine to citrulline and
C:Genetics:
A:Introns: 46/3
A:Note: single copy gene
C:Superfamily: trichohyalin; calmodulin repeat homology
C:Keywords: calcium binding; citrulline; EF hand; hair; tandem repeat
Z:49-81/Domain: calmodulin repeat homology <EP2>
F:387-851/Region: 28-residue repeats
F:886-1519/Region: 23-residue repeats

Query Match 9.3%; Score 251.5; DB 1; Length 1549;
Best Local Similarity 22.7%; Pred. No. 0.0019;
Matches 111; Conservative 108; Mismatches 194; Indels 77; Gaps 17;

QY 6 QPEAGPQAQRPQAAPVAEAGPSSQAPRKPEGQAQRTAQGALRDVSELSRQLE 65
Db 521 QREERKRRERQYLEKVELOEELQREKREKREKQ-----YLEKVELQEE 574
QY 66 ILSTYCDVNNQGGPGEAGQEPAPEDAESKRTYVARNGEPEPTPVVYGEKPSKGD 125

Query Match 9.1%; Score 247; DB 2; Length 1001;
Best Local Similarity 21.9%; Pred. No. 0.0018;
Matches 142; Conservative 118; Mismatches 212; Indels 176; Gaps 30;

QY 23 PAVEA3-----GPGSSQA-PRKPGGAQARTAQSCALRDVSEELSRLE 64
DB 324 PAVEA3EEBEOHGGRTGTNSVGSNUSISMSISASSQSSVNSLFDASDDKS-ELD 382
QY 65 DILSTYCYVNNQ-----GPGEDGAQGEPAEPDAEKSRITYVARNEPEPTPVVYGEKSPS 120
DB 383 MNEGDHTVNSNSVHLKPEENYO-EGDPR-----TRAPAQPPQVSRHKSXY 432
QY 121 KGPNTBEIROSDEVDORHRRPOEKKAKGLGKEIT-----LLMOTNLTSPEEKLAAL 176
DB 433 RAREHPATITASLV-----TRQCEHQSDESEKQSGYKRRROKQKQMLTENKIKAE 488
QY 177 CKKYA-----ELLEFEHN--SQKQMLLOKKQKQSOLOVE-----KOHLRGESHKAV 219
DB 489 MDEHRLRLDKOLETQNNFAEMEKLIKQASMEKEAKVMANERKKPQCHIQAQOKKEL 548
QY 220 LARSKLESICRELQRNRSIKBEGVOR-----AREEE-----KRKEVTSHEQV--TLNDIQL 270
DB 549 --NSPFSQKREYKLRKEQLKEELNENQSTPKKEQEWLSKQENTOHFOABEEANILRR 606
QY 271 QMB-----OHN-----ERNSKLRQENMELAEKLIKLEQVYELREEH 306
DB 607 QRYVLEECRPFKRMMLGRHNEQDLVRELNKRTQKLEHAMLLRQHESQLELFRH 666
QY 307 IDKVPKHDLOQOLVDAKQQAQEMLEKAEERHOREKDFLLKEAVSORMCHLMKQOETH 366
DB 667 LNTIQK---MRCBLL--RLQHOTELTNQLEYNRRERELRKEVMEVRRQPPKSLSKBLQ 721
QY 367 LKQOL---ALYTEKEEFONTLSKSEVFTTFQKQEMKTKIKKLEKETTMYRSMES 422
DB 722 IKQOFQDCKIQRYKALRNHLLE-----TTPKSEHAKVRLK--BEQTRKLAILEQ 774
QY 423 SNKALLEMAEKTVR-----DKELEGL 444
DB 775 YDHSINEMLSQALRLDEAQBACVOLKQLOLEBLNAYQSKIKQMAEQAQDRELREL 834
QY 445 QVXI-----QLEKLCHALQTERND-----LNKEVQDLSAGGGS-----LTD 482
DB 835 EQRVSLRALLQEKIEEMALQALQTERIRSLRQARIEAPDSESMLRGLFSNNVLSN 894
QY 483 SGPRRPEG-PGAQPS-SRVVTRAP--CYPGPAPSTEASG---QTGPQ 523
DB 895 LSPEAFESHYPGASSWSHNPNGSGPHGHPMGGTPOAGHPMGQGPQ 942

RESULT 20
S22695
restin - human
C:Species: Homo sapiens (man)
C:Date: 04-Dec-1992 #sequence_revision 04-Dec-1992 #text_change 05-Nov-1999
C:Accession: S22695; S19853
R:Biibe, G.; Delabie, J.; Brueggan, J.; Richener, H.; Asselbergs, F.A.M.; Cerletti, N.;
EMBO J. 11, 2103-2113, 1992
A:Title: Restin: a novel intermediate filament-associated protein highly expressed in th
A:Reference number: S22695; MUID:92289675; PMID:1600942
A:Accession: S22695
A:Molecule type: mRNA
A:Residues: 1-1427 <BL>
A:Cross-references: EMBL:X64838; NID:935998; PIDN:CAA46050.1; PID:g35999
C:Keywords: cytoskeleton

Query Match 9.1%; Score 246; DB 2; Length 1427;
Best Local Similarity 24.4%; Pred. No. 0.0029;
Matches 114; Conservative 84; Mismatches 150; Indels 118; Gaps 20;

QY 117 KEPSKDPNTBEIROSDEVDGR-----DHRPQEKKAKGLGKE-----ITLMTQTLN 164
DB 770 KASSBGSKMKLRQQLAEABQKHLKLEIKQVASSKASSITRLEQGLKLTNLQENLS 829

QY 165 TLS-----TPEKLAALCKKVAELLEER-----HRNSOKOMKLLQKKQSOL----- 204
DB 830 EVSQVKETLEKELQILKEKFAAEEAAVSQVRSQKQETVYKHLQKEEQFNMLSSDLKLE 889
QY 205 -----VQEKHLRGESHKAVLARSKLSELCSELQOR-----HNRS-----LKEEG 243
DB 890 NLADWEAKFREKDE---REEQLIKAKRLENDIAEIMQSGDSSQLTMMNDELKRD 946
QY 244 VQRAEEEEKEKEVTSHPQVTLNDIQLQMEQ-HVERNSKLRQENMELAEKLIKLEQVYL 302
DB 947 VRELQKLTKANENASFLQKSIDMTVKAQSOQRAAKGHEEKLEKPLKSLDEKKMET 1006
QY 303 REEHIDKV-----FKHKD-----LQQLVDA--KLQQAQOE-----MLKEABE-RH 339
DB 1007 SHNQOQLKARYERATSETTKGHEILLQKTLTLDTKLKGARENSGLIQELEELRK 1066
QY 340 QREKDFLKEAVSORMCELMKQO-----ETHLQOALALYTEKPEEPONT--L 385
DB 1067 QADKAKAAQTAADAMQIMEQMTKETETLASLEDTKQTNALQNELDTLKENNLKNVSEL 1126
QY 386 SKSSSVFTTFQKQEMKTKIKKLEKETT-----MYRSRMSSNKA 426
DB 1127 NKSKELLTVENQKMEFPKKEIETLKQAAAQSQLSALQAEENVKLAEBLGRSRDEVTSQ 1186
QY 427 LLEMAEKTVRDKLEGLQVQKIQRLKLCRALQTERNDLNKRVQDLSA 474
DB 1187 KLE--EERSVLNNQL--LEMK-KRESKFIKQADESKASQKSIKSI 1229

RESULT 21

I38055
myosin heavy chain, perinatal skeletal muscle - human
N:Contains: myosin ATPase (EC 3.6.4.1)
C:Species: Homo sapiens (man)
C:Date: 17-May-1996 #sequence_revision 17-May-1996 #text_change 19-Apr-2002
C:Accession: I38055; JH0154; S12459; S09332; A30220; S49478
R:Jullian, E.H.; Kelly, A.M.; Pompidou, A.J.; Hoffman, R.; Schiaffino, S.; Stedman, H.H.
Eur. J. Biochem. 230, 1001-1006, 1995
A:Title: Characterization of a human perinatal myosin heavy chain transcript.
A:Reference number: I38055; MUID:95324556; PMID:7601129
A:Accession: I38055
A>Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 1-1937 <RES>
A:Cross-references: EMBL:X38133; NID:9558668; PIDN:CAA86293.1; PID:g558669
R:Karsch-Mizrachi, I.; Peghali, R.; Shows, T.B.; Leinwand, L.A.
Gene 89, 289-294, 1990
A:Title: Generation of a full-length human perinatal myosin heavy-chain-encoding cDNA.
A:Reference number: JH0154; MUID:90323631; PMID:2373371
A:Accession: JH0154
A:Molecule type: mRNA
A:Residues: 1-14, 'A', 16-859 <KAR>
A:Cross-references: GB:Y00821
A:Experimental source: skeletal muscle
R:Bober, E.
submitted to the EMBL Data Library, January 1989
A:Reference number: S12458
A:Accession: S12459
A:Molecule type: mRNA
A:Residues: 502-1071, 'N', 1073-1250, 'DGG', 1253-1376, 'NT', 1379-1913, 'D', 1915-1937 <BOB>
A:Cross-references: EMBL:X51592; NID:929465; PIDN:CAA35941.1; PID:g29466
R:Bober, E.; Buchberger-Seidl, A.; Braun, T.; Singh, S.; Goedde, H.W.; Arnold, H.H.
Eur. J. Biochem. 189, 55-65, 1990
A:Title: Identification of three developmentally controlled isoforms of human myosin he.
A:Reference number: S09331; MUID:90235862; PMID:1691980
A:Accession: S09332
A:Molecule type: mRNA
A:Residues: 502-547, 'X', 549-617, 'X', 619-687, 'X', 689-757, 'X', 759-827, 'X', 829-897, 'X', 899-1376, 'NT', 1379-1386, 'X', 1388-1456, 'X', 1458-1526, 'X', 1528-1596, 'X', 1598-1666, 'X', 1668-1
A:Cross-references: EMBL:X51592
R:Peghali, R.; Leinwand, L.A.

C:Species: Mus musculus (house mouse)
 C>Date: 20-Feb-1995 #sequence_revision 20-Feb-1995 #text_change 05-Nov-1999
 C:Accession: S49461, S59599
 R:Julien, S.; Luc, M.; Francois, C.
 Submitted to the EMBL Data Library, October 1994
 A:Description: Cloning and sequencing of the murine SCP1 cDNA.
 A:Reference number: S49461
 A:Accession: S49461
 A>Status: Preliminary
 A:Molecule type: mRNA
 A:Residues: 1-993 <JUL>
 A:Cross-references: EMBL:238118; NID:gl360015; PIDN:CAA86262.1; PID:g558603
 Biochem. J.; Martin, L.; Cuzin, F.; Rassoulzadegan, M.
 R:Saigh, J.; Mathis, Acta 1263, 258-260, 1995
 A:Title: cDNA sequence of the murine synaptonemal complex protein 1 (SCP1).
 A:Reference number: S59599; MUID:96004899; PMID:7548215
 A:Accession: S59599
 A>Status: Preliminary
 A:Molecule type: mRNA
 A:Residues: 1-993 <SAG>
 A:Cross-references: EMBL:238118; NID:gl360015; PIDN:CAA86262.1; PID:g558603

Query Match 9.0%; Score 243.5; DB 2; Length 993;
 Best Local Similarity 22.9%; Pred. No. 0.0026;
 Matches 124; Conservative 91; Mismatches 183; Indels 143; Gaps 23;

QY 34 QAPRPEGAQARTAGSALRDVSELSKQLSD-----ILSTVCVD 73
 DB 132 QENRRIIEAQRKAIQIE--LQFENEKVSLSKEEIQENKDLIKENNATTHWNLKFTC-- 187
 QY 74 NNQGGGEGDGAQEPAPEDAKSTYVARNGEPEPTVVVYGEKPSKGDPTNEEIRQSD 133
 DB 188 -----ARSAKTKYVERBETRGVYVDLSNIEKILAF-----EELVQA 229
 QY 134 EVG-----DRDHRFPQ--BKAKAGLGKEITLLMQLTNTLSTPPEKLAALCKYVBL 183
 DB 230 ENARLEMFKLKEDHEKIQHLEEEYQKEVNNKENQVSELLTQSAEKENKM-----KOLTF 285
 QY 184 LEEHNSQKQM-----KLQKQSQVQKHLRGHSHKAVLARSKLESICRELRH--N 236
 DB 286 LEESDKANLEETKLOQENLKELSEKQHL-----TSELDIKMSHORSSTQ 335
 QY 237 RSLKEE-----GVQARREBEKKEVTSHPQV-----TLNDIQLOMBQHN 276
 DB 336 KALEBDQIATNTISQLETVKAEQMEELNKAHTTHSFVTELTAKTCTTLELLRTBOORL 395
 QY 277 ERNS-----KLQENKELAERLKKLIEQVELREBEHIDKVPKHDLQQLQVDAKLOQA 328
 DB 396 EKNEPQLKLITVELQKKNEL--EMTKFNKNEVELBELKNILAE--DQKLDEK-KQV 450
 QY 329 QMLKAEERHQREKDFLL-----KEAVESQWCELMKQOETH-LKQQLALYTE-KFEEFPQ 382
 DB 451 EKLAELQEKQO-ELTFLLETRKEVEHDLQEQVTVTKTSBQHYLKQVEEMKTELEKEKL 509
 QY 383 NT-----LSKSSSEVPTTFKQEMKMTKKIKLEKETMY 416
 DB 510 NTELTASCDMLLENKKFTVQASDVALEKQHQEDINCKQGBERLQKLENLEKEMHL 569
 QY 417 RSRWESSNKALLEMAEE-KTVDRKELE---GLQVKIQLEKLCALQATERNDLNKRVQDL 472
 DB 570 RDELESVRKEFTQQGDEVKCKLDKSEENARSIECEVLKKEKQMKILSKCNLKKQVENK 629
 QY 473 S 473
 DB 630 S 630

RESULT 24

A56734

ribosome receptor, 180k - dog

C:Species: Canis lupus familiaris (dog)

C>Date: 08-Jul-1995 #sequence_revision 03-Aug-1995 #text_change 08-Oct-1999

C:Accession: A56734

R:Wanker, E.B.; Sun, Y.; Savitz, A.J.; Meyer, D.I.
 J. Cell Biol. 130, 29-39, 1995
 A:Title: Functional characterization of the 180-kD ribosome receptor in vivo.
 A:Reference number: A56734; MUID:95310363; PMID:7790375
 A:Accession: A56734
 A>Status: preliminary; not compared with conceptual translation
 A:Molecule type: mRNA
 A:Residues: 1-1534 <WAN>
 A:Cross-references: GB:X87224; NID:g984113; PIDN:CAA60676.1; PID:g984114
 C:Keywords: endoplasmic reticulum; membrane protein; protein biosynthesis
 F:198-743/Region: 10-residue repeats (N-Q-G-K-K-A-E-G-A-P)

Query Match 9.0%; Score 242; DB 2; Length 1534;
 Best Local Similarity 20.4%; Pred. No. 0.0046;
 Matches 153; Conservative 114; Mismatches 222; Indels 260; Gaps 30;

QY 1 KSSPCQPEPAGER-----GAQERPSQAPAVEAREGSSQAP 36
 DB 710 KKEGSPNQKKVDASANQSKRAESAPITQGGKADWVQSEAPKQZAPKCKSGSKGKGP 769
 QY 37 RXP-----EQAQAR-----TAQSGALRDVSEE----- 58
 DB 770 GPPDSPLPLPYKTLVSTVGVNFEQAQRLIILLSEKAGVTDTHKATQKQDPVAL 829
 QY 59 LSRQLED---ILSTVCVNNQGGGEGDGAQEPAPEDAKSRT---YVARNGEPEPTPV 112
 DB 830 LKROLEEKELIAT-----EODAAVAKSKLREVNKELAAEKAKA 869
 QY 113 VTGEPEPSK---GDPNTEETIROQSDVGDHRDHPQEKKKAK-----G 151
 DB 870 AAGEAKVKQLVARQBITAVQARIRASYREHVKEVQQLQKIRTLQBLENGENTQLAR 929
 QY 152 LKQETILLQTLN-----TLSTPEKLAALCKYAEELLEHRNSQKQMLQK 199
 DB 930 LQENSIILDALNQATQVSEKQNTLEAKLQKELSKVSKELVEKSEAAQEQOKKALET 989
 QY 200 KQSQVQKQHLRGHSHKAVLA-RSKLESICREL---QRNRSLSKEGVQVQARBEERKX 255
 DB 990 KTAALKQVLQASHKESSEALQKLDSEVSELCSRTSHASLRAD-AEKARQOQQQMA 1048
 QY 256 EYTSHFQVTLNDIQLOMBQHNERSKL---RQBNWELAERLKK---LIEQVELRE----- 304
 DB 1049 ELHSLKQSSRAEVKSKSELSGLHGLKEARANSQMKRIRSTEALLEGAGQARDTQDAQ 1108
 QY 305 -----EHIDKVPKHDLQQLQVDAKLOQAEMLKEAERHQ-----REKDFLLKEAVES- 353
 DB 1109 ASRAEHQARL---KELESQV--WCLEKBATLKEAVQKVKNNDLREKNWKAMEALASA 1163
 QY 354 QWCELMKQOETHLKQO-----LAL-----YTE----- 376
 DB 1164 ERACEKRLSRUTQAKBSEKOLSLTRACTKEALLALLPALSSAPQSYTEMLQELREKGP 1223
 QY 377 -----KFEFPQNTLSKSSEVTTFKQEMKMTKKIKK-LEKETT 414
 DB 1224 ELKQRPADTPSSDLASKLREAEETQNNLQAECDQVRTILAEETGLMKOLQKSVBEEQ 1283
 QY 415 MYRSWESSNKALLEMAEEKTVRDKKELEGQVKIQ-----BLEKLCR 456
 DB 1284 V---WKAKVATTEBHLQKSRVTVYKHLDEIVKLEGELESSEQVREHTSHLEAELEKHA 1339
 QY 457 ALQTERNDLNKRVQDLGAGGQCSLTDSGPERRRPPGQAGAPSS-----PRVTEAPCY-- 508
 DB 1340 AASAEQCSYAKEV---AGLEQLLLESQSLDAKSEAKQSNELALVRQQLSEKSHVE 1395
 QY 509 -----PGAPSTASGQTGPQEBTSARA 530
 DB 1396 DDVAGSPAAPAP-----QDPVELKA 1417

RESULT 25

A56539

giantin - human

N:Alternate names: macrogolglin

A:Reference number: AS9234; PMID:96291845; PMID:8663323
A:Accession: A59234
A:Status: preliminary; not compared with conceptual translation
A:Molecule type: mRNA
A:Residues: 1-1931 <NIK>
A:Cross-references: GB:U53862; NID:g1289513; PIDN:AAC59912.1; PID:g1289514
C:Superfamily: myosin heavy chain; myosin motor domain homology
P:81-761/Domain: myosin motor domain homology <WMO>

Query Match 8.9%; Score 240; DB 2; Length 1931;
Best Local Similarity 21.7%; Pred.No. 0.0071;
Matches 110; Conservative 92; Mismatches 182; Indels 122; Gaps 17;

QY 40 EGAQARTAGSALRDVSELSRQLIEDILSTYCDNNQGGPDGAGQGPABEDAEKST 99
DB 1090 EDEQALAAQ--LQKKLKEQLARIEEL-----EELEAERTGAKVEKLS 1132
QY 100 YVARNGEPPTPVYVGEKPSKGDENTSEIROSDEVGRDHRHPQKKKARGLGEITLL 159
DB 1133 ELIQ--ELEET-----SERLEEAGGATSVOL-----ELNKKREAFQGRDLEEAFL 1178
QY 160 MDTLNTLSTPFEEKLAALCKY---AELEEHNSQKQMKLLQKQSOLVQEKHLRGEH 215
DB 1179 -----QHEATAAALRKCHADSVAEISEQLDNNQVRKQKLEKSELKLELDVSSNM 1230
QY 216 SKAVLARSKLSLCR-----ELQHRNSLKE----- 241
DB 1231 EQLIKAKANLEKNCSTEDDNNHRNKLEESQVTTDLSQRAKLQTNSELSRQLEKE 1290
QY 242 -----EGVQARSEERKREKVTSH-FQVTLNDIQLQWQHNRNSKLK 283
DB 1291 AFINQLMRGKLYTQQLDLKRLQSEAKAKALAHALQSAQHDCLLEQVSEE---M 1346
QY 284 QENMRLAERLKL--IEQYELREHIDKVFHKDLQ--QLVDAKLQQAQ-- 330
DB 1347 EKAELQRALSTANSEVAQWRTYB-TDAIQRTELEBAKKLAQLQSEAEVAZNAK 1405
QY 331 --MLKEASRPHQREKDFLLKEAVEGRCYCELMKQOETHKQOLAYITPKPFFQNTLSKS 388
DB 1406 CSSLEKTKHRLQNEITLADYERSNAAALDKQRNFDKILSEWKPFESQTELEAS 1465
QY 389 SVFVTFQKEMSKTKIKLEKETTYVRSWESSNKALLEMASEKTVRDKLEGLQVKI 448
DB 1466 QKEASLSLTEL----FKLNAVESELEHLYFKRNKNLQHEISDLT---EQLGASQSKI 1518
QY 449 QRLEKLCRALQTERNDLNKRVODLSA 474
DB 1519 HELEKVRKQLDAEKLELQAALEEA 1544

RESULT 27
S28261
centromere protein E - human
N:Alternate names: centromere 312K protein; kinesin-related protein CENP-E
C:Species: Homo sapiens [man]
C:Date: 03-Mar-1994 #sequence_revision 03-Mar-1994 #text_change 19-Jan-2001
C:Accession: S28261
R:Yen, T.J.; Li, G.; Schaar, B.T.; Szilak, I.; Cleveland, D.W.
Nature 359, 536-539, 1992
A:Title: CENP-E is a putative kinetochore motor that accumulates just before mitosis.
A:Reference number: S28261; PMID:93024922; PMID:1408971
A:Accession: S28261
A:Molecule type: mRNA
A:Residues: 1-2663 <YEN>
A:Cross-references: EMBL:Z15005; NID:G29864; PIDN:CAA78727.1; PID:G29865
C:Genetics:
A:Gene: GDB:CENPE
A:Cross-references: GDB:361164; OMIM:117143
A:Map position: 4q24-q25
C:Superfamily: centromere protein E; kinesin motor domain homology
C:Keywords: ATP; coiled coil; microtubule binding; mitosis; nucleotide binding; P-loop
P:7-335/Domain: kinesin motor domain homology <KMT>
P:86-93/Region: nucleotide-binding motif A (P-loop)

F:486-2183/Domain: coiled coil #status predicted <COI>
F:92/binding site: Arp (lys) #status predicted

Query Match 8.9%; Score 240; DB 1; Length 2663;
Best Local Similarity 20.2%; Pred. No. 0.01;
Matches 130; Conservative 101; Mismatches 176; Indels 238; Gaps 25;

QY 55 VGEELSRQLEDLSTYCVNNQGGPGEDGAGQ-----EPAEPEDAESSTYVA-----102
DB 987 ISSEVSRLNH-----MEENTGETDEFQWGWGIDKQQLAANTQTLDVQVNEII 1039
QY 103 -----RNGEPBPTFVYGEKPSKGD-----PNTFEIR-QSDFEVGRDH 140
DB 1040 EQQRKIFSLIORELQWLESVIAEKEQLKTDLKENIENTIENQEBELRGLDELKKQOE 1099
QY 141 RPOQKKA-KGLGKEITLLMTLSTPBEKL-----173
DB 1100 IVAQENHAKKEG-----LSYCDRLAEVBEKLSKQQLQEQQLLVQVEMSEKOK 1155
QY 174 -----AALCKKYAELIEHRNSOKQWKLQKQKQSVQVE 207
DB 1156 KINEIENLXNELNKLTLHEMTERLELAQKLXNENYEVKSITKERVKLQKQSPETE 1215
QY 208 KDHLAG-----EHSKAV--LARSKLSLC-----RELQHN 236
DB 1216 RDLRGYIREIATGELQYSELKIAHILKEHQETIDELRSVSEKTAQINTODLEKSH 1275
QY 237 RSLKEGVQARHEEKKKEV--TSHFVTVNDIQLOQEQHNENS-----KLRQENNELA 290
DB 1276 TKLOEE-IPVLHEQELLFNVKVSQETQETWNELELLETSQSTTTLARIEMERLRLN 1334
QY 291 ERLKXLIQELUREHIDKV-----FKHKLQQLQVLD--AKLQAAQ-----329
DB 1335 EKFQSQSEIKSLTKERDNLKTIKALEVQKDLQKHEIRLAKIQSQSQKQEQSLNMXE 1394
QY 330 -----EML--YEAEERHOREKDFLLKXAVESQRMCEL 359
DB 1395 KDNETTKIVSEMEQFKPKDSALLRIEIMGLSKRLQESHDEMS--VAKENKDLQLQEV 1453
QY 360 MKQOETHLKOOL-----ALVTEKPEFO-----NTLSKSSSEVFTTFKQ 397
DB 1454 LQESDQLKENIKVIAVHLTEBELVAKHCLKEQSETINELVNLSEKTELSTLQKQ 1513
QY 398 ---EMEKTKKIKKL-EKETM--YRSRWESSNKALLEMAEKTVDKELGLQVKI--448
DB 1514 LEAINDKLQKIQIETVEEQLNKIQISEVQENVELKQFKHEHKAQDSALQSTESKMLE 1573
QY 449 -----ORLEKLCRALQTERNDLKNRVODLSA 474
DB 1574 LTNRLQESQEBZIQIMKEEKERVQBALQIERDQLENKEIVVA 1618

RESULT 28
A45973
trichohyalin - human
C:Species: Homo sapiens (man)
C:Date: 03-May-1994 #sequence_revision 01-Mar-1996 #text_change 22-Jun-1999
C:Accession: A45973
R:Lee, S.C.; Kim, I.G.; Marekov, L.N.; O'Keefe, E.J.; Parry, D.A.D.; Steinert, P.M.
J. Biol. Chem. 268, 12164-12176, 1993
A:Title: The structure of human trichohyalin. Potential multiple roles as a functional
ed (cross-linking) protein.
A:Reference number: A45973; MUID:93280194; PMID:7685034
A:Accession: A45973
A:Molecule type: DNA
A:Residues: 1-1898 <LEE>
A:Cross-references: GB:L09190; NID:G292835; PIDN:AA65582.1; PID:G292836
A:Note: authors translated the codon AGG for residue 1714 as Pro
C:Comment: Trichohyalin is a protein of the medulla of the hair and of the inner root sh
Covalent modifications to this protein include conversion of arginine to citrulline and
C:Genetics:
A:Gene: GDB:THH
A:Cross-references: GDB:136223; OMIM:190370

A:Map position: lq21-lq21
C:Superfamily: trichohyalin; calmodulin repeat homology
C:Keywords: calcium binding; citrulline; hair; tandem repeat
F:49-81/Domain: calmodulin repeat homology <EF2>

Query Match 8.9%; Score 239.5; DB 1; Length 1898;
Best Local Similarity 22.3%; Pred. No. 0.0074;
Matches 125; Conservative 102; Mismatches 231; Indels 103; Gaps 21;

QY 8 RAGPEGAQERPSQAAPAVEAGPGSSQAPRPGGAQAARTAGSALRDVSELSQLEIDIL 67
DB 262 EEPQORBELQSEBQLKLERQELRRERQEEQQQLRRQQLRRQKQERREQQE--319
QY 68 STYCVNNQGGPGEDGAGCEPAEPEDAESSTYVARNCEPEPTPVYGEKPSKGPNTPE 127
DB 320 -----ERREQQERREQQE--ERREQQERREQQERREQQERREQQE---Q 362
QY 128 ETRQSDVGRD-----HRRPQKKKAKGLGKEITLLMTLSTLSTPEKLAACKKYABLL 184
DB 363 QLRREQEERREQQERREQQERREQQERREQQERREQQERREQQERREQQERRE 409
QY 185 EEH--RNSQKQMLQKQKOSQVQKDLHGEHSHKAVLARSKLSLCLRELQHRNLSKEE 242
DB 410 REQQLRRQQLRRQQLRRQQLRRQQLRRQQLRRQQLRRQQLRRQQLRRQQLRR 464
QY 243 GVQARAREEE--KKEVTSHFVTVNDIQLOQEQHNENSKL-----RQENNELAERLKK 295
DB 465 RLKREQEERDRLKKEEBETHEQEERKQQLRRQEEERREERMLKLEEEERREQQERRE 524
QY 296 LT--ROYELREHIDKVKHDLQQL-----VDKLLQQAQEMLKAEERHQ 340
DB 525 QLRREQEERREQQERREQQERREQQERREQQERREQQERREQQERREQQERRE 584
QY 341 R-----EKQPLLEAVESSQKELMKQOETHLQKQALYTERKEPEFQNTLSKSEVFT 393
DB 585 RLKREQEERDRLKKE--ERRQQLKKEQEERLEQLK--RREVERLEQS--ERRDELK 640
QY 394 TKQEMKMTKKIKKLEKETTMYRSWESSNKALLEMAEKTVDKELGLQVKIQRLK 453
DB 641 REEPERREHLLKSEQE-----ERRHQLREQEERREQQERREQQERREQQERRE 693
QY 454 LCRALQTERNDLKNRVQDLSAGQSGSLTDSQPERPE-----GPGAAQSPSPRVTAPCY 508
DB 694 -----EHEERREQLAEERQEQARERIKSRPKWQWOLESEADARQSKVILEAP--743
QY 509 PGAPSTEASQGTGPOEPTSR 529
DB 744 -----QAGRAEAPQEQEKR 758

RESULT 29
G02520
plectin - human
C:Species: Homo sapiens (man)
C:Date: 21-Dec-1996 #sequence_revision 06-Jun-1997 #text_change 20-Aug-1999
R:McLean, W.H.I.; Smith, P.J.D.
submitted to the EMBL Data Library, March 1996
A:Accession: G02520
A:Reference number: H01385
A:Status: Preliminary; translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 1-4574 <MCL>
A:Cross-references: EMBL:U53204; NID:G1477645; PIDN:AAB05427.1; PID:G1477646
C:Genetics:
A:Gene: PLECL
C:Superfamily: plectin; alpha-actinin actin-binding domain homology; ribosomal protein
F:68-283/Domain: alpha-actinin actin-binding domain homology <ACT>
Query Match 8.8%; Score 239; DB 2; Length 4574;
Best Local Similarity 24.9%; Pred. No. 0.019;
Matches 150; Conservative 92; Mismatches 235; Indels 126; Gaps 26;

171 EKALCKYKAYBALLEHRSNOM--KLQKQSQVQKQKHLRGHSHKAVLARSLSL 228
 2428 QEAARL-----RLAERDLAQORALAEKLMKRM-QAVQAEATLKAEE----- 2468
 229 CRELQHNESLKEGVORAREEERK-----EVTSHFQVTL--NDIQLOMBOHNRNSK 281
 2469 -ALIQOQKELAQEQARRLOADKEQVQVQVETQGFQRTLAERQRLKESAEAR-LK 2526
 282 LRQENN-----ELAEKLLKLEQVYELREHIDKVPKHLQOQVDAKLQQAQ---E 330
 2527 LRMAESRAQAAEEDAQFRQAEBIGELKHLRTLQATQKVTQVTLQIQRQSQDQAE 2586
 331 MLKBAERHQRKDFLLKEAVSQRCELMK--QOETHLQKQALYKTEKEFEFQNTLSKS 388
 2587 RLFEATAELEKEKLUQKAKULQKSEKQVQVQVQVQVQVQVQVQVQVQVQVQV 2646
 389 SEVFTTFKQEMKNTK-----KIKLEKETMYRSWESSNKALLEMAEKTVDKEL 441
 2647 RFI-----EQEKAKLEQFQDEKAKAQVQEQEQEQEQEQEQEQEQEQEQEQEQ 2702
 442 -EGL-----QVQKQLEKLC-----RALQTERNDLNRVODLS-----ACQCSL 480
 2703 EGVREKQEQELQEQ 2762
 481 TDSGPRRP--EGPGQA 496
 2763 TKALPNGRDALDPSMEA 2780
 RESULT 32
 T13030
 microtubule binding protein D-CLIP-190 - fruit fly (Drosophila melanogaster)
 C:Species: Drosophila melanogaster
 C>Date: 13-Aug-1999 #sequence_revision 13-Aug-1999 #text_change 17-Nov-2000
 C:Accession: T13030
 R:Lantz, V.A.; Miller, K.G.
 J. Cell Biol. 140, 897-910, 1998
 A>Title: A class IV unconventional myosin is associated with a homologue of a microtubul
 A:Reference number: Z17588; MUID:98139549; PMID:9472041
 A:Accession: T13030
 A>Status: preliminary; translated from GB/EMBL/DBJ
 A:Molecule type: mRNA
 A:Residues: 1-1690 <LAN>
 A:Cross-references: EMBL:AF041382; NID:92773362; PID:92773363; PIDN:AAB96783.1
 A:Experimental source: strain Oregon R
 C:Genetics:
 A:Cross-references: FlyBase:FBgn0020503
 C:Keywords: cytoskeleton
 Query Match 8.8%; Score 237.5; DB 2; Length 1690;
 Best Local Similarity 23.1%; Pred. No. 0.0079;
 Matches 126; Conservative 104; Mismatches 178; Indels 137; Gaps 24;
 33 SQAPKPEGAQARTAQSGALRDVSELSROLEDILSTYCVVDNNGGPGEDGAQEPAPPE 92
 352 SSTPVKILLATPKSQFS--WQDLREKQVQVKEKLVREDDLK-----E 392
 93 DAERSRTYVARNGPPTPVVYGEKPSKGPDPNTEIRQS-DEV---GDRHRPQ-EKK 147
 393 DAQNOALQLOQVNIQVIELKARIVELESALGNRKKTTELQCSIDRAQFCQDBLNAQSQVYKE 452
 148 KAKGLGKEITLMQ---TINTLSTP-----EKLALCKK-----YALL 184
 453 KIHLESKITKLVSATPSLOSLPDLPSDDGALQAEIAQIQEQTTCQKEVESRIABQL 512
 185 BEHNSQKMLKQK-----QSQVQKQKHLRGHSHKAVLARSLSLCHRELQHRNLSLX 240
 513 BEQQLRWVYLNQIATLQSELVSKDEAL-----EKFSLECCGIENLRLEL-----ALLK 564
 241 EGVORAREE-----EKKKEV-----TSHFQVTLNDI 268
 565 ENEKQAEQAEAEFRKLAESVFLRLSSELQNLKATSDSLESERVNKSDECEILOTEV 624

269 QLOMBOHNRNSK-----ROENMELAE-----RLKCLIEQVYELREHIDKVPKHKDL 316
 625 RNRDEQIRELQOQDEVTTLQNVQKADSSALDDMLRLOK--EGTEEKSTLEKTI----- 676
 317 QOQVADAKLQQAQEMLEKEAERHQRKDFLLKEAVSQRMC-----ELMKQOETHL 367
 677 EKELVQIK-EQAATLQDKQLEKQISD--LKQLAEQKLVREKTENAINQIQLESKES-I 732
 368 KQALALYTEKEFEFQNTLSKSSEVFTTFKQ-----EMEKMTKKIKLEKETMYRSRW 420
 733 EQALQKQNELEDFOKQSSSEVHLQEIKAQNTOKDLELVESGSLKLLQOOLEKTLGH 792
 421 ESSNKALLEMAEETKV-----RDKELEGVQVQKLEKLCRALQTERNDLNRVQDLSAGG 476
 793 EKLQAALBELKKEKETIIKEKEQELQQLQSKSAESSEALKVVQVQLEQLOQ--QAAASGE 850
 477 QGSLT 481
 851 EGSKT 855
 RESULT 33
 I48153
 myosin heavy chain beta, cardiac muscle [similarity] - golden hamster
 C:Species: Mesocricetus auratus (golden hamster)
 C>Date: 02-Jul-1996 #sequence_revision 02-Jul-1996 #text_change 02-Feb-2001
 C:Accession: I48153; A28298
 R:Wang, R.; Sole, M.J.; Cukerman, E.; Liew, C.C.
 J. Mol. Cell Cardiol. 26, 1155-1165, 1994
 A>Title: Characterization and nucleotide sequence of the cardiac alpha-myosin heavy cha
 A:Reference number: I48153; MUID:95115033; PMID:7815459
 A:Accession: I48153
 A>Status: preliminary; translated from GB/EMBL/DBJ
 A:Molecule type: DNA
 A:Residues: 1-1934 <RES>
 A:Cross-references: GB:L12104; NID:9402371; PIDN:AAA62313.1; PID:9402372
 R:Jandreski, M.A.; Sole, M.J.; Liew, C.C.
 Nucleic Acids Res. 16, 4737, 1988
 A>Title: Sequence of cDNA encoding the Syrian hamster cardiac beta-myosin heavy chain.
 A:Reference number: A28298; MUID:88247788; PMID:3380703
 A:Accession: A28298
 A:Molecule type: mRNA
 A:Residues: 962-965, 'E', 967-980, 'E', 981-985, 'Q', 987-1007, 'A', 1009, 'E', 1011, 'RKT', 1015-1
 536, 'L', 1538-1555, 'K', 1557-1934 <JAN>
 A:Cross-references: GB:I07273; NID:949640; PIDN:CAA30256.1; PID:949641
 A>Note: the authors translated the codon GIG for residue 1504 as Ieu
 C:Genetics:
 A:Introns: 66/3; 114/3; 167/1; 176/2; 212/3; 243/3; 265/1; 298/1; 332/3; 379/1; 418/3; 4
 23/3; 1389/2; 1450/3; 1506/1; 1547/3; 1650/3; 1718/3; 1760/3; 1852/3; 1884/3; 1929/3
 C:Superfamily: myosin heavy chain; myosin motor domain homology
 C:Keywords: actin binding; ATP; cardiac muscle; coiled coil; heart; muscle; nucleotide I
 F;87-765/Domain: myosin motor domain homology <MMOT>
 F;177-184/Region: nucleotide-binding motif A (P-loop)
 Query Match 8.8%; Score 237; DB 2; Length 1934;
 Best Local Similarity 23.3%; Pred. No. 0.0096;
 Matches 129; Conservative 88; Mismatches 208; Indels 128; Gaps 20;
 12 EGAQRPQAAVAEAGGSGQAQKPKGAQARTAQSGALRDVSELSROLEDILSTYC 71
 1034 EGSLEQEKVRMDLE-----RAXKLEGLDKLQES--IMDLND-KQDLDEKLKKD 1083
 72 VDNNGGPGEDGAQEPAPPEA-----EKSRTYVAR-----NGEPPTPVVYGEKBP 119
 1084 FELN-----ALNARIDEQALGSLQKLKELQARIELEEELEAERTAKVEKL 1135
 120 SKGDPNTEIRQS-DEVG-----DRHRRPQKKAQKGLKEITLLMTLTLSTPEE 171
 1136 SDSLSELEISELEAGGATSVQIEMNKKGAEAFQMRDLDEATL-----QHEA 1186
 172 KLAALCKY-----AELLBEHNSQKMLKQKQSQVQKQKHLRGHSHKAVLARSLSL 227

Db 1187 TAAALRKHADSVAEELGEIDNLRQVQKLEKSEFGLDVTNMEQIIKAKANLEK 1246
Qy 228 LCR-----ELQRH-----RSIKS-----241
Db 1247 MCRTLEPDQMEHRSKAEETQSYNDLTSQRAKLTQNGELSRQLDKEALISQLTGKLT 1306
Qy 242 -----EGVQAREREKKEVTSFQVTLNDLQLOMEQHNENSK-----LRQENME 288
Db 1307 YTOOEDLKRQLEBEVKAQNTLAHALQASRHDCLLREQYEESTEAKAELOCVLSKANSE 1366
Qy 289 LASELKKLIEQYELREBEHIDKVFHKDLOQLOLVDAKQAOE-----MLKAEERHQR 341
Db 1367 VAQRTKYETDAIOITELSEA--KKLAGLQDA--EEAVEAVNAKSSLEKTKHQLN 1422
Qy 342 EXDFLKEAVESORMCELMKQOETHLKKQALYTEKEPEEFQNTLSKSEVFTTFKQMEK 401
Db 1423 EIEDLAVDVERSAAAAALDKRQFDFKILAEWKQYEESSQSELESSQKEARSLSLTEL-- 1480
Qy 402 MTKIKKLEKETTMYRSWESSNKALLEMAZEKTVRDKLEGLQVKIORLEKLCRALQTE 461
Db 1481 --FKKNAYESLHLETFKKNKNLOEISDLT---EQLGSTGKSITHELEKIRKQLEAE 1535
Qy 462 RNDLKNRVQDLA 474
Db 1536 KMELOQALREAEA 1548
RESULT 34
152300
gi|152300|g|giantin - human
N:Alternate names: gcp372
C:Species: Homo sapiens (man)
C:Date: 01-Nov-1996 #sequence_revision 01-Nov-1996 #text_change 20-Aug-1999
C:Accession: I52300
R:Shoda, M.; Misumi, Y.; Fujiwara, T.; Nishioka, M.; Ikehara, Y.
Biochem. Biophys. Res. Commun. 205, 1399-1408, 1994
A:Title: Molecular cloning and sequence analysis of a human 372-kDa protein localized in
A:Reference number: I52300; MUID:95100974; PMID:7802676
C:Accession: I52300
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 1-3225 <RS>
A:Cross-references: GB:D25542; NID:9662389; PIDN:BA05025.1; PID:9808869
C:Superfamily: Giantin

Query Match 8.8%; Score 236.5; DB 2; Length 3225;
Best Local Similarity 22.8%; Pred. No. 0.017;
Matches 119; Conservative 99; Mismatches 178; Indels 125; Gaps 18;
Qy 52 LADVSELSRQLEDILSTYCVNNOGPGEDGAQCEPAEDPAEDAKSRITYVARNGEPTTP 111
Db 2248 LQGENKLLSQLETRHLY-----HSSQNELAKLESELKSL-----KQQLTD 2289
Qy 112 VYVGEKPSKGPNTPEIRQSDVGDVDRHRRPQE-----KKKAGLGKEITLLMQT 162
Db 2290 LNSLEKCKEQGNLEGIHQEADIQNSKSYEQLETDQASRLTSLAEHINMKQK 2349
Qy 163 LNTLSTPEKLAALCKKVAELREHNSQOMKLLQKQSQLVQKQHLRGEHSHKAVLAR 222
Db 2350 IISLLSGKEE--AIQVAIAELRQOH---DKETKELENLLSOEEENIVLEENKAV--- 2401
Qy 223 SKLESICRELQRHNSLKEEGYQARAREBEKEKETS-----HFQVT 264
Db 2402 DKTNGLMETL-----ATIKENIQQAQDLSFVKSMSLONDRDRIVGYQQLREHLSII 2457
Qy 265 LNDIQLOMEQHNERNSKLROE-----NMELARLKKLIEQYELREHIDKVPKH 313
Db 2458 LEKDQLIQEAAE--NNKLEETIRGLRSHMDLNSNAKLDABLI-QY---REDLQVITI 2512
Qy 314 KDLQO--QLVDKLLQAOEM-----LKAEERH-----QREKDFLLKEAVE 352
Db 2513 KDSQQLLEVLQQLQKLENYAKLSEKLSESEANEDLRRSFNALQEEKODLSKETES 2572

Qy 353 SORMCELMKQOETHLQO--LALYTEKEPEEFQNTLSKSEVFTTFKQEM-----399
Db 2573 LKVSISQLTQVTAQOEGTGLYHAQLVKVEEVRHLSALFSSQKRIAELEELVQVQ 2632
Qy 400 -----EYMTKKIKKLEKETTMYRSWESSN-----KALLEMAEKTVRDKEL 441
Db 2633 KEAAKKGVBIEKDKLKKLKHLDHAGINMETAEERVAELARDLVEMEOKLLAVTKEN 2692
Qy 442 EQLQVKIORLEKLCRALQTERNDLKNRVQDLSAGGQSLTD 482
Db 2693 KGLTAQIQSGRSMGSLQNRDHANEELDELCKKYDASLKE 2733
RESULT 35
A70387
Conserved hypothetical protein aq_1006 - Aquifex aeolicus
C:Species: Aquifex aeolicus
C:Date: 08-May-1998 #sequence_revision 08-May-1998 #text_change 02-Feb-2001
C:Accession: A70387
R:Decker, G.; Warren, P.V.; Gaasterland, T.; Young, W.G.; Lenox, A.L.; Graham, D.E.; O
V.
Nature 392, 353-358, 1998
A:Title: The complete genome of the hyperthermophilic bacterium Aquifex aeolicus.
A:Reference number: A70300; MUID:98196666; PMID:9537320
C:Accession: A70387
A:Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-978 <AQF>
A:Cross-references: GB:AE000718; NID:92983504; PIDN:AA07092.1; PID:92983515; GB:AB0006
A:Experimental source: strain V75
C:Genetics:
C:Gene: aq_1006
C:Superfamily: Archaeoglobus fulgidus conserved hypothetical protein AF1032

Query Match 8.7%; Score 236; DB 2; Length 978;
Best Local Similarity 21.1%; Pred. No. 0.0052;
Matches 107; Conservative 103; Mismatches 180; Indels 118; Gaps 17;
Qy 14 AQERPOAPVABAGPGSSQAPRKEGAQAQAQSGALRDVSEELSRQLEDILSTYCVQ 73
Db 314 AQELAR---IEAKEKFKSEKEREKLEHRLKQLQKIEILKELUS-QUSSLK-----363
Qy 74 NNQGGPGEDGAQCEPAEDPAEDAKSRITYVARNGEPTTP--VYVGEKPSKGPNTPEIRQ 131
Db 364 -----EKERYEQAKQEPEDLSERVEKGGKLVATEEKELEKE 402
Qy 132 --SDEVDGRDRHRRPOEKKAAGKIGKITLLMQTLNTLSTPEKLAALCKKVAELBEHN 189
Db 403 LFSEE-----EYTSLKMKERLIVELQRLKELKEGQLENLTQYKKEKKVHEK 452
Qy 190 SQQMKLLQKQSQLVQEKDHLRGEHSHKAVLSKLSLCRELQRH--RSLKEEGVQRA 247
Db 453 VLNELKELEK---ELKEREELHYAHVAVSYLSFGDTCVCGGIYRGKALENVDAEGISEL 509
Qy 248 REEEERKRYTSHFQVTL-----NDIQLOMSOHNERNNSKLROENME-LAEELKLLIQ 299
Db 510 KHAKELKEBEREDITTLKLVYAQKINSLEKEBEKLENEVEELKEITPENLKERIKGL-BE 568
Qy 300 YELREHID-----KVFPHKDLQOOLVDAKQAOA-----EMLKAEERHQRKDFLLKE 349
Db 569 LRLEKLEHKLKRYKALSDROKKEBAQKLAQKHAQTELELKE-----KIREKSLVKE 624
Qy 350 AVESQRCMLMKQOETHLQOALYTEKPEEFQNTLSKSEVFTTFKQEMQMTKKL 409
Db 625 FKELYRV-----ERLEDYSELSKEEINYNINSKLQIEIEKEKLLK 665
Qy 410 EKETTMYRSWESS-----NKALLEMAEKTVRDKELFEG-----443
Db 666 FEELSSRSKSLGELSALNESINSLEERKEKLEKLANIYEVAKSPREVELYLGDKEAE 725
Qy 444 LOVKIORLEKLCRALQTERNDLKNRVQD 471
Db 726 LERKIKFEESFQSLKAKKSIEEKLKE 753

A;Residues: 1-1939 <MAT>
A;Cross-references: DDBJ:D00943; NID:G219523; PIDN:RAA00791.1; PID:G219524
A;Accession: B46762
A;Status: nucleic acid sequence not shown
A;Molecule type: mRNA
R;Epp, T.A.; Dixon, I.M.C.; Wang, H.Y.; Sole, M.J.; Liew, C.C.
Genomics 18, 505-509, 1993
A;Title: Structural organization of the human cardiac alpha-myosin heavy chain gene (MYH)
A;Reference number: A49354; MUID:94140346; PMID:8307559
A;Accession: A49354
A;Status: nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 1-87, 'Q', 89-573, 'Q', 575-607, 'A', 609-743, 'T', 745-789, 'M', 791-1013, 'V', 1015-1019
1939 <PPP>
A;Cross-references: GB:220556; NID:G297023; PIDN:CAA9675.1; PID:G297024
R;Brand, N.J.; Dabhadre, K.; Yacoub, M.; Barton, P.J.R.
Biochem. Biophys. Res. Commun. 179, 1255-1258, 1991
A;Title: Determination of the 5' exon structure of the human cardiac alpha-myosin heavy
A;Reference number: S18830; MUID:92028859; PMID:1930170
A;Accession: S18830
A;Status: preliminary
A;Molecule type: mRNA
A;Residues: 1-32 <BRA>
A;Cross-references: EMBL:X56181; NID:G28318; PIDN:CAA39642.1; PID:G28319
R;Yamauchi-Takahara, K.; Sole, M.J.; Liew, J.; Ing, D.; Liew, C.C.
Proc. Natl. Acad. Sci. U.S.A. 86, 3504-3508, 1989
A;Title: Characterization of human cardiac myosin heavy chain genes.
A;Reference number: A94224; MUID:89264452; PMID:2726733
A;Accession: B32562
A;Molecule type: DNA
A;Residues: 1-87, 'Q', 89-177, 1551-1732, 'E', 1734-1736, 'T', 1738-1762, 'D', 1764-1848 <YAL>
R;Yamauchi-Takahara, K.; Sole, M.J.; Liew, J.; Ing, D.; Liew, C.C.
Proc. Natl. Acad. Sci. U.S.A. 86, 7416-7417, 1989
A;Reference number: A94226
A;Contents: erratum
A;Accession: B33835
A;Molecule type: DNA
A;Residues: 1-87, 'Q', 89-177, 1551-1732, 'E', 1734-1736, 'T', 1738-1762, 'D', 1764-1939 <YAZ>
R;Saez, L.J.; Gianola, K.M.; McNally, E.M.; Peghail, R.; Eddy, R.; Shows, T.B.; Reinwand
Nucleic Acids Res. 15, 5443-5459, 1987
A;Title: Human cardiac myosin heavy chain genes and their linkage in the genome.
A;Reference number: A93669; MUID:87260010; PMID:3037493
A;Accession: B27858
A;Molecule type: DNA
A;Residues: 1-3, 'S', 5-10, 'T', 12, 14-67 <SAE>
R;Kurabayashi, M.; Tsuchimochi, H.; Komuro, I.; Takaku, F.; Yazaki, Y.
J. Clin. Invest. 82, 524-531, 1988
A;Title: Molecular cloning and characterization of human cardiac alpha- and beta-form myosin
human atrium.
A;Reference number: A92770; MUID:88299163; PMID:2969919
A;Accession: A28908
A;Molecule type: mRNA
A;Residues: 1407-1532, 'N', 1534-1539, 'M', 1541-1576, 'NV', 1579-1704, 'EQ', 1707-1762, 'D', 1764-1764
A;Cross-references: GB:M21684; NID:G189006; PIDN:AAA36344.1; PID:G386972
A;Experimental source: fetal heart
C;Genetics:
A;Gene: GDB:MYH6
A;Cross-references: GDB:120214; OMIM:160710
A;Map position: 14q11.2-14q13
C;Superfamily: myosin heavy chain; myosin motor domain homology
C;Keywords: actin binding; ATP; cardiac muscle; coiled coil; heart; hydrolase; methylation
F:88-768/Domain: myosin motor domain homology <MMOT>
F:178-185/Region: nucleotide-binding motif A (P-loop)
F:549-586/Region: actin binding #status predicted
F:657-679/Region: actin binding #status predicted
F:841-1939/Domain: coiled coil #status predicted <COI>
F:841-1939/Region: S2
F:1282-1939/Region: light meromyosin
F:129/Modified site: N6,N6,N6-trimethyllysine (lys) #status predicted
F:184/Binding site: ATP (lys) #status predicted
F:697, 707/Active site: Cys #status predicted

Query Match	8.6%; Score 233.5; DB 1; Length 1939;
Best Local Similarity	24.1%; Pred. No. 0.013;
Matches 145; Conservative	83; Mismatches 191; Indels 183; Gaps 25;
Qy	12 EGAQRPSSAAPAVAEAGPGSSQAPRKPQGAQARTAQSGALRDVSELSQLEDILSTYC 71
Db	1037 EGSLEQEKVMDLE-----RAKKLEGDKLQTQES--IMDLEND-KLQLEKLLKKE 1086
Qy	72 VDNNOGPP-----GEDGAQGEPAEPEDAESTYVARNGEPEPTPVVYG 115
Db	1087 FDIQOQNSKIETDQVIALQLOKKLXENQARIELE-BELEAERTAK-----V 1134
Qy	116 EKBPSSGDPNTEBIROS-DRVG-----DRDHRRPOEKKKAGLGEITLLMOTLNTLS 1674
Db	1135 EKLRSDLSRELEISERLEBAGATSVQIEMNKKREAFQGERDLEAFL----- 1185
Qy	168 TPEKJLAALCKY-----AELLEHRNSOKMILLQKQSQOLVQKHDLRGHSHKAVILARS 223
Db	1186 QHEATAAALRKHGADSVAEIGEQIDMLQRVKQKLEKESFPLELDVTSNMQEIINAKA 1245
Qy	224 KLESLSCH-----ELQRNRSLSKRGVQVAR----- 248
Db	1246 NLEKVSRTLEDQANHYVKLEEARSLNDTTQRAKLQTEGSELRSQLEKEALISOLTR 1305
Qy	249 -----EEERKREKVTSH-PQVTLNDIQLQWQOINER----- 278
Db	1306 GKLSYTTQMBDKQLEEEGKAKNALAHQSAHDCDLLEQVYEETEAQAELOQVLSK 1365
Qy	279 -NSKLR-----QENWELAEARLKKLIEQVELREEHIDKV-----FKHK-- 314
Db	1366 ANSEVAQCKTYETDAIQRTTELEEAQKKLAQRLQDAEEAVANAKCSLEKTKHRLQN 1425
Qy	315 DLQQLVDAKLQQAQEMKLEABERHRQEKDFLLKEAVESQRMCLMQOETHLQOOLALY 374
Db	1426 EIEDLWYDVERSNA--AAALDKQQRNFQKILAE--WKQKYBESQSELSQSEARSLS 1480
Qy	375 TEKTEEPQNTLSKGSFVPTTFKQEMKQNTVKI-----KKLEKET 413
Db	1481 TELF-KLKNAYEBSLEHLETFKRENKQLOBEISDLTEQLGSGKNVHELEKVRKOLEVEK 1539
Qy	414 TMVSRVSSGNKALLEMAAEKTVKDEKLGLOVKIQLEKLC-----RALQTERNDLNKVR 469
Db	1540 LELOSALEEA-EASLEHBEQKILR-AQLEFNQIKABETKRLAENDEEMEQAKRN-HQRV 1595
Qy	470 QD 471
Db	1596 VD 1597

RESULT 42

B85431
trichohyalin like protein [imported] - Arabidopsis thaliana
C:Species: Arabidopsis thaliana (mouse-ear cress)
C:Date: 16-Feb-2001 #sequence_revision 16-Feb-2001 #text_change 16-Feb-2001
C:Accession: B85431
R:anonymous, The European Union Arabidopsis Genome Sequencing Consortium, The Col
Nature 402, 769-777, 1999
A:Title: Sequence and analysis of chromosome 4 of the plant Arabidopsis thaliana.
A:Reference number: B85001; MJND:20083488; PMID:10617198
A:Accession: B85431
A:Status: Preliminary
A:Molecule type: DNA
A:Residues: 1-1432 <STO>
A:Cross-references: GE:NC_001268; NID:G2770600; PIDN:CAB80318.1; GSPDB:GN00140
C:Genetics:
A:Gene: AT4g36520
A:Map position: 4

Query Match 8.6%; Score 232.5; DB 2; Length 1432;
Best Local Similarity 23.0%; Pred. No. 0.011;

24 AVEAEGPGSSOAPRKPEGAQARTAQSGALRDVSEELSRQLEDILSTYCVDDNNQGGPGEDG 83

Db 605 SVELNGNGKMKMR-----SQSETKLNEPLKRMEEB-TRIKEARLR-----EEN 647
QY 84 AQCEPAEPDAKSRITYVARNPEPTPVVYGEKEPSKDPNTBEIRQSDVEGDRDHRP 143
Db 648 DRERVAVAEKAEKRLKAALQOEKEREKINGEAREKAEENRAVEAREKAE-----698
QY 144 QEKKAAGLGLKEITLLMOTLNTLTSTPEEKLAALCKKYAELLBEHRNSKOMKLLQKQSQ 203
Db 699 QEKOM-----KE-----QOELEQLKEAFKEKEENRMRAPFALEQEKERR 739
QY 204 L--VQEKD-----HLRGEHSAVL-----ARSKLSLCRELQHRNLSKEGVQAR-----EE 251
Db 740 INEAREKEENERRIKAREKAELEQRLKATLEQEKERQIKERQEREENERRAKEVLEQA 799
QY 252 EKKKEVTSHPVTLNDIOLQMEQHNRNSKLQENNELAERLKLQYELREEHIDKVP 311
Db 800 ENERKUEALEQENSRRLKETREKENKKRLREALLEKEKRLLEAFERAB--IERL 857
QY 312 KHRDLOQLVDAKLOQA--QEMLKEAERHQ-----REKDFLKEAVESORMCE 358
Db 858 K-EDLSQEMRMRLQEAQERLHRENQEHQENRKOHEVSGESDEKERDACEMEKTC 916
QY 359 LMKQO--ETHLKOQLALYKPEFQNTLS-----KSSEVFTTFQKQEMKMTKKIKLEK- 411
Db 917 TTKEARHGEQSSNESLSDLEENESIDNVSNKQKEEECTRORESMSATCPWKVPEKT 976
QY 412 -----ETTVMSRWES-----SNKALLEMAEBKTVRDKEGLEQVKI 448
Db 977 LKDSQKEGTNEMDADTLPERNEETPRLENGGCGNQNGESGEESTVTENIIG-----1031
QY 449 ORLEKLCRALOTERN--DLNKGVDLSAGGGSITD-----SGPERRPBGCAQAPSSPRVT 503
Db 1032 GKLEQKSKNETSKDASVLKRVGLKTEVERLEDVVGVGRDQNPPE-----ESKAPKTS 1087

RESULT 43
A23767
Myosin heavy chain, fast skeletal muscle - human (fragment)
C:Species: Homo sapiens (man)
C:Date: 05-Jun-1988 #sequence_revision 05-Jun-1988 #text_change 20-Jun-2000
C:Accession: A23767
R:Saez, L.; Leinwand, L.A.
Nucleic Acids Res 14, 2951-2969, 1986
A:Title: Characterization of diverse forms of myosin heavy chain expressed in adult human skeletal muscle
A:Reference number: A93616; MUID:86176778; PMID:241254
A:Accession: A23767
A:Molecule type: mRNA
A:Residues: 1-876 <SAE>
A:Cross-references: GB:X03740; NID:g34839; PIDN:CAA27380.1; PID:g1335218
A:Superfamily: myosin heavy chain; myosin motor domain homology
C:Keywords: actin binding; ATP; coiled coil; muscle; skeletal muscle

Query Match 8.6%; Score 232; DB 2; Length 876;
Best Local Similarity 20.8%; Pred. No. 0.0069;
Matches 110; Conservative 88; Mismatches 168; Indels 162; Gaps 17;

QY 38 KPEGAQARTAQSGALRDVSELSQLELDILSTYCVNNQGGPGEDGAQGEPAPEDAEKS 97
Db 34 KIEBQALGMO--LQKKIKELQAKIEL-----EEIEAERTSAKAEKL 76
QY 98 RTYVARNGEPTPVVYGEKPSKDPNTEFI--RQSDVEGDRDHRPQEKKAKGLGKE 155
Db 77 RSDLSR-----ELEBISERLBEAVGATSTQIEMNKKRE-----AE 111
QY 156 ITLLMOTLNTLTSTPEKLAALCKK-----YALLEEHNSSQKVKLLQKQSQLVQEKDH 210
Db 112 FQMRDLEEAATLQHEATAATLKKHADSVAELGEQIDNLRVQKLEKSEKMEKNEIDD 171
QY 211 LRGEHSAVLARSKLESICREHQ-----RHNR 238
Db 172 LASNETVSKAGNLEKWCRALEDQLSEIKTKEEQRLINDLTAQRLAQNQVEYSRQ 231

QY 239 LKE-----EGVQARBEERKEKEVTSH-FQVTLNDIOLQMEQHNER 278
Db 232 LDKQDTLETOLSRGKQAFQTOIBELKRLQEBEIRKASALAHALQSSRHDCDLLREQYEE 291
QY 279 -----NSKLR-----QENNELAERLKLQYELREEHIDKVPKH 313
Db 292 QEAKAELORAMSKANSEVAQWRTYETDAIORTSELEBAKKLQALQDAEHV-----345
QY 314 KDLQOQLVDAKLOQAQEMLKEAERHQBKDFLKEAVESORMCELMKQOETHLQKQAL 373
Db 346 -----EAVNAKAS-----LEKTKQRLQNEVEDLMIDVERTNAACAALDKQTNFKILAE 396
QY 374 YTEKPEEPQNTLSKSSVFTTFQKQEMKMTKKIKLEKETTMYRSRWESSNKALLEMAEB 433
Db 397 WKQCKEE-----THAVLESFQKESRSLSTELFKI-----KNAYESLDQLETLEKRE 442
QY 434 KTVRDKELEGL--QV-----KIQRIEKLQALQOTERNDLNKRVDLSA 474
Db 443 KNNLQEQISDLSLQTEQIAEGGKRIHELEKIKKQVEQKSELQALAEAEAA 490

RESULT 44
S06006
Myosin beta heavy chain, cardiac muscle [similarity] - rat
N:Contains: myosin ATPase (EC 3.6.4.1)
C:Species: Rattus norvegicus (Norway rat)
C:Date: 31-Dec-1993 #sequence_revision 31-Dec-1993 #text_change 19-Apr-2002
C:Accession: S06006; S07536; I67441; A02989
R:Kraat, R.; Bravo-Zehnder, M.; Taylor, D.A.; Leinwand, L.A.
Nucleic Acids Res 17, 7529-7530, 1989
A:Title: Complete nucleotide sequence of full length cDNA for rat beta cardiac myosin b
A:Reference number: S06006; MUID:90016823; PMID:2798112
A:Accession: S06006
A:Status: translation not shown
A:Molecule type: mRNA
A:Residues: 1-1935 <KRA>
A:Cross-references: EMBL:X15939; NID:G56656; PIDN:CAA34065.1; PID:G56657
R:McNally, E.M.; Kraat, R.; Bravo-Zehnder, M.; Taylor, D.A.; Leinwand, L.A.
J. Mol. Biol. 210, 665-671, 1989
A:Title: Full-length rat alpha and beta cardiac myosin heavy chain sequences. Comparison
A:Reference number: S07535; MUID:90333919; PMID:2614840
A:Accession: S07536
A:Status: nucleic acid sequence not shown
A:Molecule type: mRNA
A:Residues: 1-950, 'RK', 953-1935 <MCN>
R:Mahdavi, V.; Lompre, A.M.; Chambers, A.P.; Nadal-Ginard, B.
Eur. Heart J. 5, 181-191, 1984
A:Title: Cardiac myosin heavy chain isozymic transitions during development and under p
A:Reference number: I53305; MUID:85179510; PMID:6241892
A:Accession: I67441
A:Status: translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 1871-1935 <RES>
A:Cross-references: GB:M32698; NID:G205598; PIDN:AAA41659.1; PID:G205599
R:Mahdavi, V.; Periasamy, M.; Nadal-Ginard, B.
Nature 297, 659-664, 1982
A:Title: Molecular characterization of two myosin heavy chain genes expressed in the ad
A:Reference number: A02988; MUID:82220036; PMID:7045682
A:Accession: A02989
A:Molecule type: mRNA
A:Residues: 1524-1528, 'V', 1530, 'R', 1532-1730, 'H', 1732-1783, 'K', 1785-1850, 'N', 1852-1857,
A:Cross-references: GB:J00752; NID:G205577; PIDN:AAA41654.1; PID:G205578
C:Superfamily: myosin heavy chain; myosin motor domain homology
C:Keywords: actin binding; ATP; cardiac muscle; coiled coil; heart; hydrolase; methylat
F:88-766/Domain: myosin motor domain homology <MYOT>
F:178-185/Region: nucleotide-binding motif A (P-loop)
F:548-585/Region: actin binding #status predicted
F:655-677/Region: actin binding #status predicted
F:839-1935/Domain: coiled coil #status predicted <COI>
F:839-1279/Region: 82
F:1280-1935/Region: light meromyosin
F:129/Modified site: N6,N6-trimethyllysine (lys) #status predicted
F:184/Binding site: ATP (Lys) #status predicted

F:695,705/Active site: Cys #status predicted

Query Match 8.6%; Score 232; DB 1; Length 1935;

Best Local Similarity 22.3%; Pred. No. 0.016;

Matches 116; Conservative 81; Mismatches 189; Indels 134; Gaps 16;

QY 25 VBAEGPGSCAPKPKPGQAARTAQSGALRDVSEELSRQLEDILSYCVDDNQGPGGEDGA 84

DB 1094 IDEQALGSQLKUKELQARI-----EELSEEL----- 1123

QY 85 QGRPAEPDAKSRITVARGPEPTPVVYGKPKSGKDPNTEIROSDEVDGRDHRPPQ 144

DB 1124 -AERTARAKVEKLRSDLSRELEB-----ISRLIEAGGATSVQI-----EMKKREA 1169

QY 145 EKKKAGLGKELTLLMOTLNTLSTPEKLAALCKV-----AELLSEHNSOKOMKLLQKK 200

DB 1170 EFQKWERDLEAATL-----QHEATAALKKHADSVAELGQIDNLRVQKLEKE 1220

QY 201 QSQLOVKDHLGEGHUSKAVLARSKLESICR-----EL 232

DB 1221 KSEFKLELDVTSNMEQIKAKANLEKMCRTLEDQMEHRSKABEFTORSVNDLTRQAKL 1280

QY 233 QRHN-----RSLKE-----EGVORAREEKKKEVTSH-FQVTLNDI 268

DB 1281 QTENGSELDKEALISQTRGKLYTQQLDLKRLQLEEVKAKNALAHALQSAHDC 1340

QY 269 QLOMEQHNRNSK-----LQENWELAEKLLIEQYELREZHIDKVFHKDLQQLV 321

DB 1341 DLLRSQYEBETAKAELORVLISKANSVAQMRKTYETDAIQRTSELSEA--KKLAQRLQ 1398

QY 322 DAKLOQAQ-----MLKEAERHQRKDFLLKEAVESQRMCELMKQOETHLKOQLALY 374

DB 1399 DA--BEAVAVNAKSSLEKTKHRLQNEHDLVDVDSNAAAALDKQKQNFKILVEW 1456

QY 375 TKPBFQNTLSKSSEVPTTFQEMKEMTKIKLEKETTMYRSRWSSNKALLEMAEK 434

DB 1457 KORYEESQSELSQKARSISTEL-----FKLNAYEESLREHLETFKRNKNLQEEISDL 1512

QY 435 TVRDELEGLQVKIQKLEKLCALQTERNDLNKRQDLISA 474

DB 1513 T---EQLSGTGSIHLEKIKRQLEAKLELQSALEAEA 1549

RESULT 45

A59236

embryonic muscle myosin heavy chain - sea squirt (Halocynthia roretzi)

C:Species: Halocynthia roretzi

C>Date: 19-May-2000 #sequence_revision 19-May-2000 #text_change 08-Sep-2000

C:Accession: A59236

R:Araki, I.

Submitted to GenBank, February 1999

A:Reference number: A59236

A:Status: preliminary; not compared with conceptual translation

A:Molecule type: mRNA

A:Residues: 1-1927 <ARA>

A:Cross-references: GB:D45163; NID:q1197167; PIDN:BA08111.1; PID:q1197168

A:Experimental source: clone lib lambda gtl1; dev stage tailbud embryo

C:Genetics:

C:Superfamily: myosin heavy chain; myosin motor domain homology

F:89-766/Domain: myosin motor domain homology <MWO>

Query Match 8.6%; Score 231.5; DB 2; Length 1927;

Best Local Similarity 21.0%; Pred. No. 0.016;

Matches 111; Conservative 96; Mismatches 196; Indels 125; Gaps 16;

QY 8 EAGPGGAQERSQAAPAV-----EAGPGSSQAQPKPGAQARTAQSGALRDVSELS 60

DB 1363 KANAFAQWRNKRYETDAIQRTTELEAKKLLATLQAEAEQVEATQAKASLQRTKRLQ 1422

QY 61 RQLRDLILSYCVDDNQGPGGEDGACQPAEPEDAERKRTYVARGPEPTPVVYGEKPS 120

DB 1423 GLEEDLTI-----DLERSNSAAAA--DKKQR 1447

QY 121 KGPNTTEIROSDEVDGRDHRPPQKKKAKGLGKEITLLM-----QTLNLTSTPEKLAAL 176

DB 1448 NFDKVLAKSKQBB--EIOVELEQAQKEARGSLSTELFKONSYEESLDALETVKRENKQL 1505

QY 177 CKYAEILLSEHNSOKOMKLLQKXQOLVOEKDHLR-----GEHSAVLARSKL 225

DB 1506 QSEIADLTDLQEGGKRSHELEKAKRTLEHNEIQAALREABGAISGESKVLRLQVEL 1565

QY 226 ESLCRLQHNRSLSKEGVQVRAREEEXRKE-----VTSHFQVTLND 267

DB 1566 AQIKQDPER-RLSEKSEETENTRRNQRALESMTOTLDSBSKSRQEAVERMKCKMBGDLND 1624

QY 268 IOLQMEQHNRNSKLRQENMELAEKLLISQYELREEHIDKVFHKDLQOO----- 319

DB 1625 LSIQCHATROASQSKSVKTFQAHKDLLELDVDSQHSQ-----DLQSQPAVIERE 1678

QY 320 -LVDKLOQAQEMLEAEERHQRKDFLLKEAVESQRMCELMKQOETHLKOQLALYTKF 378

DB 1679 NLIKABIDELRSALQAE-----RGRKLAETELLESSESSNLLHTQNTALINQ-----KKKL 1730

QY 379 E-BFQNTLSKSSEVPTTFQEMKEMTKIK-----KKLEKETTMYRSRWSSNKALLEM 430

DB 1731 EGELQNVQSEVEAVQEGORNAEAKKAITDAAMWAELEKKEQDI-----SSHLEKMK 1784

QY 431 ABEKTVDRKB-----LEGLQVKIQKLEKLCALQTERNDLNKR 468

DB 1785 NTEQTVKDLQRLDEABQVALLGKGGKQVQKLETRVRELENELDSEQR 1832

RESULT 46

S06005

myosin alpha heavy chain, cardiac muscle [similarity] - rat

N:Alternate names: alpha-myosin heavy chain

N:Contains: myosin ATPase (EC 3.6.4.1)

C:Species: Rattus norvegicus (Norway rat)

C>Date: 31-Dec-1993 #sequence_revision 31-Dec-1993 #text_change 19-Apr-2002

C:Accession: S06005; S07535; A20971; A02988; 153305

R:McNally, E.M.; Gianola, K.M.; Leinwand, L.A.

A:Title: Complete nucleotide sequence of full length cDNA for rat alpha cardiac myosin

A:Reference number: S06005; MUID:90016822; PMID:2798111

A:Accession: S06005

A:Status: translation not shown

A:Molecule type: mRNA

A:Residues: 1-1938 <MCN>

A:Cross-references: EMBL:X15938; NID:G56654; PIDN:CAA34064.1; PID:G56655

R:McNally, E.M.; Kraft, R.; Bravo-Zehnder, M.; Taylor, D.A.; Leinwand, L.A.

J. Mol. Biol. 210, 665-671, 1989

A:Title: Full-length rat alpha and beta cardiac myosin heavy chain sequences. Comparison

A:Reference number: S07535; MUID:90133919; PMID:2614840

A:Accession: S07535

A:Status: not compared with conceptual translation

A:Molecule type: mRNA

A:Residues: 1-1938 <MC2>

R:Mahdavi, V.; Chambers, A.P.; Nadal-Ginard, B.

Proc. Natl. Acad. Sci. U.S.A. 81, 2626-2630, 1984

A:Title: Cardiac alpha- and beta-myosin heavy chain genes are organized in tandem.

A:Reference number: A20971; MUID:84194059; PMID:6585819

A:Accession: A20971

A:Molecule type: protein

A:Residues: 1-12, 'Ap', 14-45, 'A', 47-50, 'Ap', 53-81, 'E', 83-86, 'Q', 88-109, 111-133, 'H', 135-1.

R:Mahdavi, V.; Periasamy, M.; Nadal-Ginard, B.

Nature 297, 659-664, 1982

A:Title: Molecular characterization of two myosin heavy chain genes expressed in the ad

A:Reference number: A02988; MUID:8220036; PMID:7045662

A:Accession: A02988

A:Molecule type: mRNA

A:Residues: 1512-1574, 'S', 1576-1721, 'N', 1723-1851, 'N', 1853-1869, 'N', 1871-1933, 'I', 1935-

A>Note: there are 10 or more myosin heavy chain genes in the rat, at least two of which

R:Mahdavi, V.; Lompre, A.M.; Chambers, A.P.; Nadal-Ginard, B.

Eur. Heart J. 5, 181-191, 1984

A:Title: Cardiac myosin heavy chain isozytic transitions during development and under pa
A:Reference number: 153305; MUID:85179510; PMID:6241892
A:Accession: 153305
A:Status: translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 1872-1933; 'I', 1935-1938 <RES>
A:Cross-references: GB:M2697; NID:G205596; PIDN:AAA41658.1; PID:G205597
C:Superfamily: myosin heavy chain; myosin motor domain homology
C:Keywords: actin binding; ATP; cardiac muscle; coiled coil; heart; hydrolase; methylate
P:87-767/Domain: myosin motor domain homology <MMOT>
P:177-184/Region: nucleotide-binding motif A (P-loop)
P:1548-585/Region: actin binding #status predicted
P:1658-678/Region: actin binding #status predicted
P:1840-1938/Domain: coiled coil #status predicted <COI>
P:1840-1938/Region: S2
P:1281-1938/Region: light meromyosin
P:128/Modified site: N6,N6-trimethyllysine (Lys) #status predicted
P:183/Binding site: ATP (Lys) #status predicted
P:696,706/Active site: Cys #status predicted

Query Match 8.6%; Score 231.5; DB 1; Length 1938;
Best Local Similarity 24.4%; Pred. No. 0.016;
Matches 146; Conservative 80; Mismatches 197; Indels 175; Gaps 25;
QY 12 EGAORPQAPAVAEAGSGSQAPRPEGAQARTAGSALRDVSELSQLEDLSTYC 71
DB 1036 EGSLEQEKVMDLE-----RAKRLGDLKLTQES--IMDLEND-KLQLEKLLKKE 1085
QY 72 VDNQGGPGDGAGCEPAPEDA-----EKSRTYVAR-----NGPEPTPVVYGEKEP 119
DB 1086 FDISQ-----QNSKIEDQALALQKLEKQENQARIEELEEAEARTAKVEKLR 1137
QY 120 SKGDPNTEIRQS-DEVG-----DRDHRPQKKAAGLKGKEITLLMQLTNTLSTPEE 171
DB 1138 SLDLEELSEISERLEAGGATSVQIEMNKKREAFQWRDLLEATL-----QHEA 1188
QY 172 KLAALCKKY-----AELLEHNNSOKMKLQKQSQLVQSKHURGHSHKAVLARSKLES 227
DB 1189 TAAALRKHADSVAGLQIDNLQVRQKLEKSEKSEFKLELDVDTSHMEQIIKAKANLEK 1248
QY 228 LCR-----ELQHNRSLEKEGVQAR-----KLEKETTMVR 248
DB 1249 VSRILEQANRYRVKLEAQRSLNDFPTQAKLQTEGELARQLEKEKALINQITRGKLS 1308
QY 249 -----ESEEKKEVTSH-FQVTLNDIOLQMEQHNER-----NSK 281
DB 1309 YTOQMEDLKRQLEBEGKAKNALAHQASRDHCDLLSEQYEEEMAKAEQVLSKANSE 1368
QY 282 LR-----QENNELAERLKLIEOVELREEHIDKV-----FKHK---DLQ 318
DB 1369 VAQWRTKYETDAIQRTTELEBAKKAQRLQDAEAEVAVNAKCSSLEKTKHRLQNEIED 1428
QY 319 QLVDAKQQAQEMKLEAEERHQRKDFLLKAEVSEQRMCMLKQOEHLKQALALYTEKF 378
DB 1429 LMDVVERSNA---AAAALDKQRNFDKILAE--WKQKYESSQSELESSQAEARSLSLSELF 1483
QY 379 EEFQNTLSKSSEVTFPFQEMEKMTKKI-----KKLEKETTMVR 417
DB 1484 -KLKNAYESLEHETPFQENKNI-QEISDLTEQLGSGGNVHELEKIRQLEVEKLEIQ 1542
QY 418 SRWSSNKALLEMAEKTVRQKLEGLQVKIQRLKJC---RALQTERNDLKNRVOD 471
DB 1543 SALEEA-EASLEHEEGKILR-AQLEFQIKAEIERKLAEDKBEMEQAKRNHL--RVVD 1596

RESULT 47

149464
alpha cardiac myosin heavy chain - mouse
C:Species: Mus musculus (house mouse)
C:Date: 02-Jul-1996 #sequence revision 02-Jul-1996 #text change 02-Feb-2001
C:Accession: 149464; 149463; 149462; 149461; 149604
R:Quinn-Laguer, B.K.; Kennedy, J.E.; Wei, S.J.; Beisel, K.W.
Genomics 13, 176-188, 1992

A:Title: Characterization of the allelic differences in the mouse cardiac alpha-myosin
A:Reference number: A38207; MUID:92250040; PMID:1577481
A:Accession: 149464
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 1-1938 <RES>
A:Cross-references: GB:M76601; NID:G191623; PIDN:AAA37162.1; PID:G191624
A:Accession: 149463
A:Molecule type: mRNA
A:Residues: 1-193, 'D', 195-837, 'S', 839-955, 'N', 957-1938 <RE2>
A:Cross-references: GB:M76600; NID:G191621; PIDN:AAA37161.1; PID:G191622
A:Accession: 149462
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 1-1938 <RE3>
A:Cross-references: GB:M76599; NID:G191619; PIDN:AAA37160.1; PID:G191620
A:Accession: 149461
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 1-544, 'A', 545-1938 <RE4>
A:Cross-references: GB:M76598; NID:G191617; PIDN:AAA37159.1; PID:G191618
R:Gulick, J.; Subramanian, A.; Neumann, J.; Robbins, J.
J. Biol. Chem. 266, 9180-9185, 1991
A:Title: Isolation and characterization of the mouse cardiac myosin heavy chain genes.
A:Reference number: 149604; MUID:91225025; PMID:2026617
A:Accession: 149604
A:Status: translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-67 <RES>
A:Cross-references: GB:M62404; NID:G192609; PIDN:AAA37424.1; PID:G192610
C:Superfamily: myosin heavy chain; myosin motor domain homology
C:Keywords: ATP; cardiac muscle; heart; nucleotide binding; P-loop
P:88-768/Domain: myosin motor domain homology <MMOT>
P:178-185/Region: nucleotide-binding motif A (P-loop)

Query Match 8.6%; Score 231.5; DB 2; Length 1938;
Best Local Similarity 24.2%; Pred. No. 0.016;
Matches 143; Conservative 81; Mismatches 195; Indels 173; Gaps 24;
QY 12 EGAORPQAPAVAEAGSGSQAPRPEGAQARTAGSALRDVSELSQLEDLSTYC 71
DB 1037 EGSLEQEKVMDLE-----RAKRLGDLKLTQES--IMDLEND-KLQLEKLLKKE 1086
QY 72 VDNQGGPGDGAGCEPAPEDA-----EKSRTYVAR-----NGPEPTPVVYGEKEP 119
DB 1087 FDISQ-----QNSKIEDQALALQKLEKQENQARIEELEEAEARTAKVEKLR 1138
QY 120 SKGDPNTEIRQS-DEVG-----DRDHRPQKKAAGLKGKEITLLMQLTNTLSTPEE 171
DB 1139 SLDLEELSEISERLEAGGATSVQIEMNKKREAFQWRDLLEATL-----QHEA 1189
QY 172 KLAALCKKY-----AELLEHNNSOKMKLQKQSQLVQSKHURGHSHKAVLARSKLES 227
DB 1190 TAAALRKHADSVAGLQIDNLQVRQKLEKSEKSEFKLELDVDTSHMEQIIKAKANLEK 1249
QY 228 LCR-----ELQHNRSLEKEGVQAR-----KLEKETTMVR 248
DB 1250 VSRILEQANRYRVKLEAQRSLNDFPTQAKLQTEGELARQLEKEKALISQITRGKLS 1309
QY 249 -----ESEEKKEVTSH-FQVTLNDIOLQMEQHNER-----NSK 281
DB 1310 YTOQMEDLKRQLEBEGKAKNALAHQASRDHCDLLSEQYEEEMAKAEQVLSKANSE 1369
QY 282 LR-----QENNELAERLKLIEOVELREEHIDKV-----FKHK---DLQ 318
DB 1370 VAQWRTKYETDAIQRTTELEBAKKAQRLQDAEAEVAVNAKCSSLEKTKHRLQNEIED 1429
QY 319 QLVDAKQQAQEMKLEAEERHQRKDFLLKAEVSEQRMCMLKQOEHLKQALALYTEKF 378
DB 1430 LMDVVERSNA---AAAALDKQRNFDKILAE--WKQKYESSQSELESSQAEARSLSLSELF 1484
QY 379 EEFQNTLSKSSEVTFPFQEMEKMTKKI-----KKLEKETTMVR 417

Db 1485 -RLKWAYESLEHLETFKREKNLQGEISDLTEQLGEGQNVHELEKIRKQLEVKLELQ 1543
QY 418 SRWESSNALLERAEKTVRKELGLOVKIQRLEKLC-----RALQTERNDL 465
Db 1544 SALEEA-BASLEHSGKILR-AQLSFNQIKAEIERKLAEDBEMEQAKRNHL 1593
RESULT 48
S74245
C:Species: Mus musculus (house mouse)
C:Date: 29-Jan-1998 #sequence_revision 06-Feb-1998 #text_change 24-Sep-1999
C:Accession: S74245
R:Nakagawa, O.; Fujisawa, K.; Ishizaki, T.; Saito, Y.; Nakao, K.; Narumiya, S.
PDB: 1ett, 392, 189-193, 1996
A:Title: ROCK-I and ROCK-II, two isoforms of Rho-associated coiled-coil forming protein
F:90-354/Domain: protein kinase homology <kin>
A:Reference number: S74244; MUID:96368048; PMID:8772201
A:Accession: S74245
A:Status: preliminary; nucleic acid sequence not shown
A:Molecule type: mRNA
A:Residues: 1-1388 <NAK>
A:Cross-references: EMBL:U58513; NID:g1514697; PIDN:AAC53133.1; PID:g1514698
C:Superfamily: unassigned Ser/Thr or Tyr-specific protein kinases; protein kinase C zinc
C:Keywords: phosphotransferase; serine/threonine-specific protein kinase
F:1261-1315/Domain: protein kinase C zinc-binding repeat homology <xzn>
Query Match 8.5%; Score 231; DB 2; Length 1388;
Best Local Similarity 21.5%; Pred. No. 0.012;
Matches 133; Conservative 84; Mismatches 172; Indels 230; Gaps 23;
QY 91 PEDAKSR-----TYVARNGEPE-----PTPVV--- 113
Db 316 PDDTBSKXKGLICAPLTDREVLGRNGVEIKQHPFFKNDQNWMDNIRETAAPVPEL 375
QY 114 -----YGEKEPKSGDPTSEI-----ROSDEVG 136
Db 376 SSDIDSNFDDIEDDKGDEVPFIPKAFVGNQLPIGTYFERNLLSDSPPCRENDAIQ 435
QY 137 DRDHRPQ-----KKAKGLGKEITLMTL 163
Db 436 TRKSESQGIQKVALBEHLSSEVOAKELEQCKKSINTELEKTAKELEEBEITLRKSE 495
QY 164 NTLSTPPEKLAALCKKYAEL-----LEEHRNSOK-QMKLLQKK----- 200
Db 496 STLQLEKELKALQKNAEYQKADHEADKKENLNDVNSLKQLEDLKKENQSSQISTE 555
QY 201 -----QSQLVQKZHLRGHESKAV-LARSKLES--LCRELQHRNSLKER-- 242
Db 556 KYNQLQKQLEDEANALLRTESPTAELRKTQAESKQIQLESNNRDLQDKNCLLETAKLK 615
QY 243 -----GVQARAEERKEKVTSHFPQVTLNDIQ-----LQMEQH 275
Db 616 LEKEFINLSALESERRDR---THGSEIINDLQGRISGLEBDLKTGKALLAKVELEKRL 672
QY 276 NERNKSLRQ--NWELEARKKLBOYLRREHIDKVPFKHQLQQLVD-----A 323
Db 673 QEKLTDLKEKSNBIDMTYQLVKVIQQLSLEQFEAS-----HKTTRALADRNKVIESTEA 728
QY 324 KLQQAQAEMLKE-ABERHOREK-DFILKEAVE-----SQRMELMKQOETHLKQ 369
Db 729 KSEAMKEMEKLEERSLKQKVENLLAEKRCISILDCLAQSQKLNELLKQKV-LNE 787
QY 370 QLAALYTERKEE-----FNTLSKSEVFTT-----FKQEMEKTKIKKLEKETT 415
Db 788 DVYRNTLTKEQTQKRCMLQNDLRKMQTQVNTLKMSEXQIKQENHLMEMQNLKONTE 847
QY 416 YRSRWESSNALLERAE-----KTVRDEKLEGLQVKIQRLEKLCRALQTERNDLNKR 468
Db 848 LKREQDADGQMKELQDQLEASQVSTLYKTQVRELKEENBEKTKLCKELQKKQQLQDE 907
QY 469 VQDLSAGQGGSLTDSGPER 487

Db 908 RDSLAAQLAQLAITLTKADSEQ 926
RESULT 49
T08880
NMDA receptor-binding protein yotiao - human
C:Species: Homo sapiens (man)
C:Date: 11-Jun-1999 #sequence_revision 11-Jun-1999 #text_change 21-Jul-2000
C:Accession: T08880
R:Lin, J.W.; Wyszynski, M.; Madhavan, R.; Sealock, R.; Kim, J.U.; Sheng, M.
J. Neurosci. 18, 2017-2027, 1998
A:Title: Yotiao, a novel protein of neuromuscular junction and brain that interacts wit
A:Reference number: Z16511; MUID:98151389; PMID:9482789
A:Accession: T08880
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 1-1642 <LIN>
A:Cross-references: EMBL:AF026245; NID:g2623067; PIDN:AAB86384.1; PID:g2623068
C:Genetics:
A:Map position: 7q21-22
C:Keywords: brain; cerebral cortex; coiled coil; neuromuscular junction; skeletal muscul
Query Match 8.5%; Score 230.5; DB 2; Length 1642;
Best Local Similarity 18.4%; Pred. No. 0.015;
Matches 119; Conservative 110; Mismatches 195; Indels 223; Gaps 23;
QY 31 GSSQAPRPEGAQTAQSGALRDVSEELSQLEDILSTYCVDNQ-----GGPGEQG 83
Db 161 GAQDSPTHEMMSELSA--GKQHEI-BELNRELSEMRVTYGTGLQLOLQEFEEAIAKORDG 217
QY 84 -----AQGEPAPEDEAKSRTYVARNGEPEPTPVVYGEKPSKGDPT----- 126
Db 218 ITQTATNLQARRKEDTMEFLELTESQSLQIQFOLOQASTLRNSTHSTAAADLLQ 277
QY 127 ---EBIROSDVGDHRRPQEKKAKGLGKEITLMTLNTLSTPBEKLAALCK----- 179
Db 278 AKQILTHQOQLLEQHLLEDYQKKEDFTQISFLQEKIKVYEMEQDKKVENSKZBIQ 337
QY 180 -----YAELESEHNSOKQ-----KLQKQSQLVQKZHLRGHESKAVLARS 223
Db 338 EKETIIBELNTKIEBEKNTLELKDKLTTADKLGLQELQIVQKNQIK--NMKLELTNS 395
QY 224 K-----LESCLRELQHRN--SLKBEGVQARAEERKEK----- 256
Db 396 KQKERQSSEIKQLMGTVEELQKRNHDKSQFETDIVORMEQETQRLKLEAFELDEMYGQ 455
QY 257 -----VTSH-----FQVTLNDIQLQME 273
Db 456 QIVQKKQELIRQHAQMEKTRHGENALRSYNTNEDQIKLMNVAINELNKLQ 515
QY 274 QHNRNSKLK-----QNNMELARLKKLBOYLYELREHIDKVPK-----HX 314
Db 516 DTNSQKELKEELGLILEKCALQRLQLEDLVEILSFSEQIQARQRTIAEQESKLENAHK 575
QY 315 DLQO-QLVDAKLQQAQAEMLKEAERH-----OREKDFELKEAVESQRM--- 356
Db 576 SLSTVEDLKAEIVSASERKLELKHAEVNTYKLEMLEKKNVLDNRMASSQEALE 635
QY 357 -----CELMKQOQ-----THLKQQL-----ALYTEKP 378
Db 636 RLRTQLLPFSHEBELSKLDELEIHRINIEKLNGLGHYKQIQDGLQENMSQKTIETMQF 695
QY 379 BEPQNTLSKSEVFTT-----TTFQEMEKTKIKKLEKETTYSRNSKAL 427
Db 696 EK-DNLITKQNLILSILKDLQOQSLNSKSEKTIQINELQHEILR----- 744
QY 428 LEMAEKTVRKELGLOVKIQRLEKLCRALQTERNDLNKRVDLSA 474
Db 745 -QEEKEKGTLEQVQELQKLTLEK---QMKENDLQEKFAQLEA 787
RESULT 50

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: June 7, 2004, 12:41:25 ; Search time 27 Seconds
(without alignment)

1022.118 Million cell updates/sec

Title: US-10-023-529-8

Perfect score: 2702

Sequence: 1 KSSPQPPACGEGKQERPSQ.....APSTASGQTGPQPTASARA 530

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 141681 seqs, 52070155 residues

Total number of hits satisfying chosen parameters: 141681

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 150 summaries

Database : SwissProt 42:*

pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1315	48.7	259	1 YL14 HUMAN	P40222 homo sapien
2	264.5	8.8	1084	1 MYSS_RABIT	P02562 oryctolagus
3	261.5	9.7	1407	1 TRHY_RABIT	P02562 oryctolagus
4	259	9.6	1938	1 RA50 ARCFU	O92920 archaeglob
5	259	9.6	1938	1 MYSS_CHICK	P13538 gallus gall
6	256	9.5	2871	1 DESP_HUMAN	P15924 homo sapien
7	252.5	9.3	880	1 RA50 PYRAB	O92920 mus musculus
8	251.5	9.3	1549	1 TRHY_SHEEP	P22793 ovis aries
9	251.5	9.3	1690	1 C190_DROME	O92920 mus musculus
10	249	9.2	1938	1 MYH4_RABIT	P22793 ovis aries
11	246	9.1	1427	1 REST_HUMAN	P30622 homo sapien
12	245	9.1	1937	1 MYH8_HUMAN	P13535 homo sapien
13	245	9.1	1940	1 MYH3_CHICK	P02565 gallus gall
14	244.5	9.0	1410	1 RRB1_HUMAN	O92920 mus musculus
15	243.5	9.0	993	1 SCPI_MOUSE	O62209 mus musculus
16	242.5	9.0	4473	1 PLE1_CRIGR	O91155 cricetus
17	242	9.0	1534	1 RRB1_CANFA	O92920 mus musculus
18	242	9.0	3259	1 GGB1_HUMAN	O92920 mus musculus
19	240	8.9	2663	1 CENE_HUMAN	Q14789 homo sapien
20	239.5	8.9	1898	1 TRHY_HUMAN	Q02224 homo sapien
21	239.5	8.9	1939	1 MYH1_HUMAN	Q07283 homo sapien
22	239	8.8	4684	1 PLE1_HUMAN	Q15149 homo sapien
23	238.5	8.8	4687	1 PLE1_RAT	P10427 rattus norv
24	237	8.8	1934	1 MYH7_MESAU	P13540 mesocricetu
25	237	8.8	1938	1 MYH7_HUMAN	Q08433 homo sapien
26	236.5	8.8	1939	1 MYH6_HUMAN	P13533 homo sapien
27	236	8.7	978	1 RA50_AQUAR	Q07124 aquifex aeo
28	236	8.7	1411	1 EBA1_HUMAN	Q15075 homo sapien
29	234	8.7	1935	1 MYH7_HUMAN	P12883 homo sapien
30	234	8.7	1935	1 MYH7_PIG	P79293 sus scrofa
31	233.5	8.6	845	1 SCPI_MESAU	O60563 mesocricetu
32	233	8.6	1941	1 MYH2_HUMAN	O92920 mus musculus
33	232	8.6	1935	1 MYH7_RAT	P02564 rattus norv

34	231.5	8.6	882	1 RA50_PYRFU	P58301 pyrococcus
35	231.5	8.6	1938	1 MYH6_MOUSE	Q02566 mus musculus
36	231.5	8.6	1938	1 MYH6_RAT	P02563 rattus norv
37	231	8.5	1605	1 RRB1_MOUSE	O92920 mus musculus
38	230.5	8.5	1939	1 MYH6_MESAU	P13539 mesocricetu
39	230.5	8.5	1911	1 AKAG_HUMAN	O92920 h a-kinase
40	230	8.5	1939	1 MYH4_HUMAN	O92920 mus musculus
41	229.5	8.5	790	1 EBA1_MOUSE	Q08166 mus musculus
42	229.5	8.5	2230	1 GGB1_HUMAN	Q14105 gallus gall
43	229.5	8.5	2238	1 GGB1_MOUSE	P13540 mus musculus
44	229	8.5	1976	1 MYH4_HUMAN	P35580 homo sapien
45	229	8.5	2779	1 LVA_DROME	Q08551 drosophila
46	228	8.4	1938	1 MYH3_AEQIR	P24733 aequipecten
47	228	8.4	2442	1 CEB2_HUMAN	O92920 mus musculus
48	226.5	8.4	1790	1 USO1_YEAST	O92920 mus musculus
49	226.5	8.4	1940	1 MYH3_HUMAN	P11055 homo sapien
50	226.5	8.4	2210	1 CENF_HUMAN	P11055 homo sapien
51	225.5	8.3	1940	1 MYH3_RAT	P12847 rattus norv
52	225	8.3	1919	1 INCE_HUMAN	Q08433 homo sapien
53	224	8.3	1961	1 MYH9_RAT	Q02812 rattus norv
54	224	8.3	1976	1 MYH4_BOVIN	Q27991 bos taurus
55	224	8.3	1976	1 MYH4_RAT	Q27991 bos taurus
56	223.5	8.3	1679	1 GGC2_MOUSE	O91155 cricetus
57	223	8.3	492	1 M5_STEPS	O91155 cricetus
58	221.5	8.2	997	1 SCPI_RAT	P02977 streptococ
59	221.5	8.2	1102	1 MYSC_CHICK	Q03410 rattus norv
60	221	8.2	1935	1 MYSS_CYPCA	P29616 gallus gall
61	220.5	8.2	2017	1 MYSN_DROME	O90339 cyprinus ca
62	220	8.1	1330	1 KTN1_VULVU	O99323 drosophila
63	220	8.1	2349	1 TPR_HUMAN	O97961 vulpes vulp
64	219.5	8.1	852	1 RA50_THEMA	P12270 homo sapien
65	219.5	8.1	2035	1 M18A_MOUSE	O91155 cricetus
66	219	8.1	1846	1 MYSB_RAT	O91155 cricetus
67	218.5	8.1	863	1 M1SP_TAESO	P02569 rattus norv
68	218.5	8.1	1197	1 CING_HUMAN	P35418 taenia soli
69	218.5	8.1	1742	1 MYSC_HUMAN	Q092m7 homo sapien
70	218	8.1	1364	1 KTN1_CHICK	Q03631 gallus gall
71	218	8.1	1433	1 REST_CHICK	O42184 gallus gall
72	218	8.1	3321	1 PCN2_HUMAN	O95613 homo sapien
73	217.5	8.0	967	1 KINH_LOLPE	P16113 loligo peal
74	217.5	8.0	1972	1 MYHB_RABIT	P35748 oryctolagus
75	217.5	8.0	1978	1 MYHB_CHICK	P10587 gallus gall
76	216	8.0	757	1 P1BF_HUMAN	Q8wxw3 homo sapien
77	216	8.0	1191	1 CING_MOUSE	P59242 mus musculus
78	216	8.0	1357	1 KTN1_HUMAN	O96up2 homo sapien
79	216	8.0	1583	1 GGC2_MOUSE	O81w12 homo sapien
80	215.5	8.0	729	1 GOA5_MOUSE	O92920 mus musculus
81	215.5	8.0	866	1 MYSP_SCHJA	O95870 schistosoma
82	215.5	8.0	1360	1 CING_XENLA	Q08433 homo sapien
83	215.5	8.0	1959	1 MYH9_CHICK	P4105 gallus gall
84	215.5	8.0	1960	1 MYH9_HUMAN	P35579 homo sapien
85	215.5	8.0	1972	1 MYH9_HUMAN	P35749 homo sapien
86	215	8.0	697	1 MPP1_LYCES	P93203 lycopersico
87	214.5	7.9	1721	1 ITN1_HUMAN	Q15811 homo sapien
88	214.5	7.9	2033	1 MYH1_HUMAN	O92817 homo sapien
89	214.5	7.9	2245	1 MYH1_MOUSE	P54697 dictyosteli
90	214	7.9	1286	1 SMC4_MOUSE	O92920 mus musculus
91	214	7.9	1957	1 SPOF_SCHPO	Q10411 schistosom
92	213.5	7.9	767	1 GOA1_HUMAN	Q02805 homo sapien
93	213.5	7.9	880	1 INCE_MOUSE	Q08433 homo sapien
94	213.5	7.9	1714	1 ITN1_MOUSE	Q20462 mus musculus
95	213	7.9	862	1 C5P2_MACFA	Q0852 macaca fasc
96	213	7.9	1875	1 M1P1_YEAST	O02455 saccharomyc
97	213	7.9	2054	1 M18A_HUMAN	O92614 homo sapien
98	212	7.8	1300	1 DYNA_NEUCR	Q01397 neurospora
99	211.5	7.8	1027	1 KINH_MOUSE	P33175 mus musculus
100	211.5	7.8	2116	1 MYH2_DICDI	P08799 dictyosteli
101	211	7.8	1224	1 DYNA_CHICK	P35458 gallus gall
102	211	7.8	1290	1 SMC4_XENLA	P50532 xenopus lae
103	210.5	7.8	863	1 MYSP_ECHGR	P35417 echinococcu
104	210.5	7.8	1597	1 CTRO_MOUSE	P49025 mus musculus
105	209.5	7.8	1044	1 VAF3_SCHPO	Q09857 schizosacch
106	209.5	7.8	2611	1 BPLE_MOUSE	Q912u8 mus musculus

QY 452 EKLALQTRNDLNKXVQSLDSGQSGSLTDSGPRPEGPGAGAPSSPRVTEAPCPYGA 511
DB 181 EKLALQTRNDLNKXVQSLDSGQSGSLTDSGPRPEGPGAGAPSSPRVTEAPCPYGA 240
QY 512 PSTEASQGTGQPEFTSARA 530
DB 241 PSTEASQGTGQPEFTSARA 259

RESULT 2

ID MYSS RABBIT STANDARD; PRT; 1084 AA.
AC P02552;
DT 21-JUL-1986 (Rel. 01, Created)
DT 01-JAN-1990 (Rel. 13, Last sequence update)
DE 15-MAR-2004 (Rel. 43, Last annotation update)
DE Myosin heavy chain, skeletal muscle (Fragments).
OS Oryctolagus cuniculus (Rabbit).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Lagomorpha; Leporidae; Oryctolagus.
OX NCBI_TaxID=9986;
RN [1]
RP SEQUENCE OF 1-258.
RA Capony J.-P., Eizinga M.;
RT "The amino acid sequence of A 34,000 dalton fragment from S-2 of
RT myosin.";
RL Biophys. J. 33:148A-148A (1981).
RN [2]
RP SEQUENCE OF 259-428.
RX MEDLINE=85131142; PubMed=3972832;
RA Lu R.C., Wong A.;
RT "The amino acid sequence and stability predictions of the hinge
RT region in myosin subfragment 2.";
RL J. Biol. Chem. 260:3456-3461 (1985).
RN [3]
RP SEQUENCE OF 409-1084 FROM N.A.
RX MEDLINE=87304245; PubMed=3305014;
RA Maeda K., Szakiel G., Wittinghofer A.;
RT "Characterization of cDNA coding for the complete light meromyosin
RT portion of a rabbit fast skeletal muscle myosin heavy chain.";
RL Eur. J. Biochem. 167:97-102 (1987).
CC 1- FUNCTION: Muscle contraction.
CC 1- SUBUNIT: Muscle myosin is a hexameric protein that consists of 2
CC heavy chain subunits (MHC), 2 alkali light chain subunits (MLC)
CC and 2 regulatory light chain subunits (MLC-2).
CC 1- SUBCELLULAR LOCATION: Thick filaments of the myofibrils.
CC 1- DOMAIN: The rodlike tail sequence is highly repetitive, showing
CC cycles of a 28-residue repeat pattern composed of 4 heptapeptides,
CC characteristic for alpha-helical coiled coils.
CC 1- MISCELLANEOUS: Each myosin heavy chain can be split into 1 light
CC meromyosin (LM) and 1 heavy meromyosin (HMM). It can later be
CC split further into 2 globular subfragments (S1) and 1 rod-shaped
CC subfragment (S2).
CC 1- SIMILARITY: Contains 1 myosin-like globular head domain.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb.ch/announce/>
CC or send an email to license@isb.ch).
CC
CC EMBL; X05958; CAA29391.1; -;
CC F01; A02985; A02985.
CC F01; A05280; A05280.
CC F01; S00084; S00084.
CC Myosin; Muscle protein; Coiled coil; Thick filament; Actin-binding;
CC ATP-binding; Multigene family.
CC
CC NON TER 1 1 ALPHA-HELICAL TAILPIECE (S2).
CC DOMAIN <1 >258
CC NON CONS 258 259
CC DOMAIN <259 1084 RODLIKE TAIL (S2 AND LAM DOMAINS).
CC

FT DOMAIN 455 1084 COILED COIL (POTENTIAL).
FT VARIANT 405 L -> V.
FT VARIANT 408 V -> L.
FT VARIANT 421 E -> D.
FT VARIANT 423 S -> G.
FT VARIANT 426 K -> R.
SQ SEQUENCE 1084 AA; 125488 MW; 229CFD69A61E7F0 CRC64;

Query Match 9.8%; Score 264.5; DB 1; Length 1084;
Best Local Similarity 22.7%; Pred. No. 0.00071; Indels 167; Gaps 20;
Matches 131; Conservative 85; Mismatches 194;

QY 12 EGAQRPSPQAAPVAEAGFGSSQAPRKPEGAQARTAQSGALRDVSELSRQLEDILSTYC 71
DB 193 EGSLEQEKIRMDLE-----RAKKLEB-DLKLAEQSM-DIEND-KOOLDEKLKLE 242
QY 72 VDNQGGPDED-----GAQGPAPEDAEKSRFYVARNGEPEPTVVVG 115
DB 243 FMTNLSQKTEQALMTNLRQIELEBELEAEASAKAKORSDLSRELE-----IS 296
QY 116 EKEPSKIDPTEIRQSDVGDHRRPQEKKAKGLGKEITLLMQTLNLTSTPEBKLA 175
DB 297 ERLEHAGGATSAQI-----EMNKKBAEFKMRDLLEATL-----QHEATAAA 340
QY 176 LKKY-----AELLEHRNSQKQKLLQKQSQQLVQEKHLRGEHSKAVLARSKLESLCE 231
DB 341 LRKKGADSVAEELGEQIDNLRVQKLEKESBLKMEIDDLAGNMTVTSKAGNLEKWCRT 400
QY 232 L-----QRHNRSLKE----- 241
DB 401 LEDQLSEVTKKEEHOQLNELNSAQARLHTSEGFQRQLDEKDMVSQLSRGGQFTQQ 460
QY 242 -EGVQARREBEKKEVTSH-FQVTLNDILOMEQNER-----NSKLR-- 283
DB 461 IEGLKQLSEETKAKSALAHALQSSRRDCDLLREQYEEQEAKAELORAMSKANSEVSQW 520
QY 284 -----QENNELAERLKLLEIQVELREEHIDKVFHKHLDLQQLVDAKLQQAQBMKE 334
DB 521 RYKCTDALQRTHELEAAKKLAQLQDAEHEV-----EAVNSKCA--LEK 565
QY 335 ABERHQENDFLKAEAVESQRMCELMKQOETHLKQQLALYTEKFEFQNTL-----S 386
DB 566 TKQLQNEABDLMI DVERSNATCARMDKQKRNFDKVLAEWKGHYEETOAELEASQESRS 625
QY 387 KSEVETTFK-----QEMKMTKKIKLEKETTMYRSWESSNKALLWABEKTVDKE 440
DB 626 LSTEVFKVKNAYEESLDHLETLKRNKNIQEELSDITEQLASAKHILEKVKQKIDQE 685
QY 441 LEGLOVKIQRLB-----KLCRALQTRNDLNKRV 469
DB 686 KSELQALAEAGSLSEHSGKIUR-IQLBLNQVKSSEI 721

RESULT 3

ID TRHY RABBIT STANDARD; PRT; 1407 AA.
AC P37709;
DT 01-OCT-1994 (Rel. 30, Created)
DT 01-OCT-1994 (Rel. 30, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Trichohyalin.
GN THH.
OS Oryctolagus cuniculus (Rabbit).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Lagomorpha; Leporidae; Oryctolagus.
OX NCBI_TaxID=9986;
RN [1]
RP SEQUENCE FROM N.A.
RA Fietz M.J., Rogers G.E.;
RL Submitted (DEC-1992) to the EMBL/GenBank/DBJ databases.
CC 1- FUNCTION: Intermediate filament-associated protein that associates
CC in regular arrays with keratin intermediate filaments (KIF) of the
CC inner root sheath cells of the hair follicle and the granular
CC

layer of the epidermis. It later becomes cross-linked to KIF by isodipeptide bonds. It may serve as scaffold protein, together with involucrin, in the organization of the cell envelope or even anchor the cell envelope to the KIF network. It may be involved in its own calcium-dependent postsynthetic processing during terminal differentiation.

CC -1- SUBUNIT: Homodimer (Probable).

CC -1- TISSUE SPECIFICITY: Found in the hard keratinizing tissues such as the inner root sheath (IRS) of hair follicles and medulla, and in the filiform papillae of dorsal tongue epithelium (Probable).

CC -1- DEVELOPMENTAL STAGE: Expressed during late differentiation of the epidermis.

CC -1- DOMAIN: Consists of nine domains. Domain 1 contains two EF-hand calcium-binding domains. Domains 2-4, 6, and 8 are almost entirely alpha-helical, configured as a series of peptide repeats of varying regularity, and are thought to form a single-stranded alpha-helical rod stabilized by ionic interactions. Domain 5 is the most regular and may bind KIF directly by ionic interactions. Domains 5 and 7 are less well organized and may induce folds in the molecule. Domain 9 contains the C-terminus, conserved among different species.

CC -1- PTM: Substrate of transglutaminase. Some 200 arginines are probably converted to citrullines by peptidylarginine deamidase.

CC -1- SIMILARITY: In the N-terminal section; belongs to the S-100 family.

CC -1- SIMILARITY: Contains 2 EF-hand calcium-binding domains.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

EMBL; Z19092; CA79519.1; --
PIR; S28589; S28589.
HSSP; P02633; 4ICB.
InterPro; IPR001751; CaBP_S100.
InterPro; IPR002048; EF-hand.
Pfam; PF00036; ehand; 1.
Pfam; PF01023; S_100; 1.
ProDom; PD003407; CaBP_S100; 1.
PROSITE; PS00018; EF_HAND; 1.
PROSITE; PS00303; S100_CaBP; 1.
Keratinization; Calcium-binding; Repeat; Citrullination.
DOMAIN 1 91 S-100 LIKE.
CA_BIND 22 33 EF-HAND 1 (LOW AFFINITY) (POTENTIAL).
CA_BIND 62 73 EF-HAND 2 (HIGH AFFINITY) (POTENTIAL).
SEQUENCE 1407 AA; 183781 MW; AE17D2AL59P12B7F CRC64;

Query Match 9.78; Score 261.5; DB 1; Length 1407;
Best Local Similarity 23.5%; Pred. No. 0.0012;
Matches 117; Conservative 110; Mismatches 168; Indels 103; Gaps 22;

QY 26 EAEGFGSS-----QAPKPEGAQTAQSGALRDVSELSQLELDILSTCYVDNNGG 78
DB 514 QEQPGQWQWQEQEQRNHTYAPGQSQURE--EELQR-----EKRRQE 561

QY 79 PGEDGAGEPAEPDAEKSRITYVARNCEPPTVYGEKSPSGKDPNTHIRQSDVGR 138
DB 562 RERYREEEKLQREDEKRR-----RQRERQYRELEBLRQEQELDR 604

QY 139 DHRPQEQKAKGKGTLLMQTLNTLSTPEKLAALCKKYAELLE--HNSQKQMKLL 197
DB 605 KLREEQQLQER---EELRLRQRERKLREEQQL--LRQEQELRQERKURLEEQQL 659

QY 198 QKQSQVLQVQEKDH-LRGEHSAVLARSKLSLCELRQHNRSLSKEBQVQARREE---EK 253
DB 660 REEQQLQERKURLEEQQL---LQERESEELRQERKURLEEQQLRQEQELQER 716

QY 254 RKEVTSFQVTLNDIQLOMEQHNENSKLRQENNEL-----AERLKKLQYELREEHDK 309
DB 254 RKEVTSFQVTLNDIQLOMEQHNENSKLRQENNEL-----AERLKKLQYELREEHDK 309

717 ERKLREEQQLRREEQQLRQ---ERDRKLREEQQLQSEERLRLRQERQQLRRER-DR 772
QY 310 VPKHND-----LQQVLDVDAKLQQAQMKLKEAE-----EEHQEKDFLAKEA 350
DB 773 KFEDEEQQLQREERLRLRQERKURLEEQQLQSEERLRLRQERKURLEEQQLQER 832
QY 351 VESORMC-----ELMKQEQTHLQKQALALYTEKFEFQNTLSKSEVFTTFKOE 398
DB 833 EERLRLRQERKURLEEQQLRQEQQLRQER---RKLREEQQLRQEQQL---ELRQE 885
QY 399 MEKMTKIKKL-----EKTTMYRWRWESSNKALLEMAEKTVRDELEGLQVKIQLEKL 454
DB 886 RDRKLREEQQLRQERKURLEEQQLRQERKURLEEQQLQSEERLRLRQER---KLREEQQL 942
QY 455 CRALQTE-RNDLNKRVOD 471
DB 943 LRREEQQLRREERKURLE 960

RESULT 4
RA50_ARCFU STANDARD; PRT; 886 AA.
ID O29230.
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE DNA double-strand break repair rad50 ATPase.
GN RAD50 OR AF1032.
OS Archaeoglobus fulgidus.
OC Archaea; Euryarchaeota; Archaeoglobi; Archaeoglobales;
OC Archaeoglobaceae; Archaeoglobus.
OK NCBI_TaxID=2234;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=VC-16 / DSM 4304 / ATCC 49558;
RX MEDLINE=98049343; PubMed=938475;
RA Klenk H.-P., Clayton R.A., Tomb J.-F., White O., Nelson K.E., Ketchum K.A., Dodson R.J., Gwinn M., Hickey E.K., Peterson J.D., Richardson D.L., Kerlavage A.R., Graham D.E., Kyriakides N.C., Fleischmann R.D., Quackenbush J., Lee N.H., Sutton G.G., Gill S., Kirkness E.F., Dougherty B.A., McKenney K., Adams M.D., Loftus B., Peterson S., Reich C.I., McNeil L.K., Badger J.H., Glodek A., Zhou L., Overbeek R., Gocayne J.D., Weidman J.P., McDonald L., Utterback T., Cotton M.D., Spriggs T., Artiach P., Kaine B.P., Sykes S.M., Sadow P.W., D'Andrea K.P., Bowman C., Fujii C., Garland S.A., Mason T.M., Olsen G.J., Fraser C.M., Smith H.O., Woese C.R., Venter J.C.;
RA The complete genome sequence of the hyperthermophilic, sulphate-reducing archaeon Archaeoglobus fulgidus.;
RL Nature 390:364-370(1997).
CC -1- FUNCTION: Involved in DNA double-strand break repair (DSBR). The rad50/mrel complex possesses single-strand endonuclease activity and ATP-dependent double-strand-specific exonuclease activity. Rad50 provides an ATP-dependent control of mrel by unwinding and/or repositioning DNA ends into the mrel active site (By similarity).
CC -1- SUBUNIT: Forms a complex with mrel (By similarity).
CC -1- SIMILARITY: Belongs to the SMC family. RAD50 subfamily.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

EMBL; AS001032; AAB90211.1; --
DR PIR; H69378; H69378.
DR TIGR; AF1032; --
DR HAMAP; MF 00449; -; 1.
DR InterPro; IPR003593; AAA_ATPase.
DR InterPro; IPR003439; ABC_transporter.

-1- FUNCTION: Muscle contraction. Myosin is a protein that binds to F-actin and has ATPase activity that is activated by P-actin.

-1- SUBUNIT: Muscle myosin is a hexameric protein that consists of 2 heavy chain subunits (MHC), 2 alkali light chain subunits (MLC) and 2 regulatory light chain subunits (MLC-2).

-1- SUBCELLULAR LOCATION: Thick filaments of the myofibrils.

-1- DOMAIN: The rodlike tail sequence is highly repetitive, showing cycles of a 28-residue repeat pattern composed of 4 heptapeptides, characteristic for alpha-helical coiled coils.

-1- MISCELLANEOUS: Each myosin heavy chain can be split into 1 light meromyosin (LMW) and 1 heavy meromyosin (HMM). It can later be split further into 2 globular subfragments (S1) and 1 rod-shaped subfragment (S2).

-1- SIMILARITY: Contains 1 myosin-like globular head domain.

-1- SIMILARITY: Contains 1 IQ domain.

This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (see <http://www.isb-sib.ch/announce/> or send an email to license@sb-sib.ch).

EMBL; U87231; AAB47555.1; --
EMBL; M16557; AAB48970.1; --
PDB; 2MYS; 11-JAN-97.
PDB; 1ALM; 17-DEC-97.
PDB; 1M8Q; 10-SEP-02.
InterPro; IPR000048; IQ region.
InterPro; IPR001609; myosin head.
InterPro; IPR004009; Myosin N.
InterPro; IPR002928; Myosin_tail.
Pfam; PF00612; IQ; 2.
Pfam; PF00063; myosin_head; 1.
Pfam; PF02736; Myosin_N; 1.
Pfam; PF01576; Myosin_tail; 1.
PRINTS; PR00193; MYOSINHEAVY.
ProDom; PD000355; myosin_head; 1.
SMART; SM00015; IQ; 1.
SMART; SM00242; MYSC; 1.
PROSITE; PS00096; IQ; 1.
Myosin; Muscle protein; Coiled coil; Thick filament; Actin-binding; ATP-binding; Methylation; Phosphorylation; Acetylation; Calmodulin-binding; Multigene family; 3D-structure.

FT	INIT MET	0	0	MYOSIN HEAD-LIKE.
FT	DOMAIN	1	782	IQ.
FT	DOMAIN	783	812	HINGE.
FT	DOMAIN	838	840	COILED COIL (POTENTIAL).
FT	DOMAIN	841	1338	ATP (POTENTIAL).
FT	NP BIND	179	186	ACTIN-BINDING.
FT	DOMAIN	657	679	ACTIN-BINDING.
FT	DOMAIN	759	773	ACETYLATION.
FT	MOD_RES	1	1	METHYLATION (MONO-).
FT	MOD_RES	35	35	METHYLATION (TRI-).
FT	MOD_RES	130	130	METHYLATION (TRI-).
FT	MOD_RES	551	551	METHYLATION (MONO-).
FT	MOD_RES	755	755	C -> Q (IN REF. 7 AND 8).
FT	CONFLICT	907	907	E -> F (IN REF. 1).
FT	CONFLICT	980	980	L -> D (IN REF. 5).
FT	CONFLICT	1343	1343	S -> A (IN REF. 5).
FT	CONFLICT	1545	1545	HV -> QL (IN REF. 5).
FT	CONFLICT	1796	1797	S -> A (IN REF. 5).
FT	CONFLICT	1830	1830	I -> V (IN REF. 10).
FT	CONFLICT	1863	1863	IHG -> FH (IN REF. 10).
FT	CONFLICT	1929	1931	
FT	TURN	7	8	
FT	TURN	11	12	
FT	HELEX	14	17	
FT	TURN	21	24	
FT	HELEX	25	28	
FT	TURN	34	36	
FT	STRAND	37	41	

RESULT 6
 DESP_HUMAN STANDARD; PRT; 2871 AA.
 AC P15924; O75993; Q14189; Q9UNH4;
 DT 01-APR-1990 (Rel. 14, Created)
 DT 16-OCT-2001 (Rel. 40, Last sequence update)
 DT 15-MAR-2004 (Rel. 43, Last annotation update)
 DE Desmoplakin (DP) (250/210 kDa paraneoplastic pemphigus antigen).
 GN DSP.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A. (ISOFORM DPI).
 RC TISSUE=Fore skin;
 RX MEDLINE=92115697; PubMed=1731325;
 RA Virata M.L.A., Wagner R.M., Parry D.A.D., Green K.J.;
 RT "Molecular structure of the human desmoplakin I and II amino
 terminus.";
 RL Proc. Natl. Acad. Sci. U.S.A. 89:544-548(1992).
 RN [2]
 RP SEQUENCE FROM N.A. (ISOFORM DPI).
 RC PHILLIPS S.;
 RX Submitted (OCT-1998) to the EMBL/GenBank/DBJ databases.
 RN [3]
 RP SEQUENCE OF 1120-2871 FROM N.A. (ISOFORM DPI).
 RC TISSUE=Poreskin;
 RX MEDLINE=90153880; PubMed=1689290;
 RA Green K.J., Parry D.A.D., Steinert P.M., Virata M.L.A., Wagner R.M.,
 RA Angst B.D., Nilles L.A.;
 RT "Structure of the human desmoplakins. Implications for function in
 the desmosomal plaque.";
 RL J. Biol. Chem. 265:2603-2612(1990).
 RN [4]
 RP ERATUM.
 RX MEDLINE=90361712; PubMed=2391353;
 RA Green K.J., Parry D.A.D., Steinert P.M., Virata M.L.A., Wagner R.M.,
 RA Angst B.D., Nilles L.A.;
 RL J. Biol. Chem. 265:11406-11407(1990).
 RN [5]
 RP SEQUENCE OF 2854-2871 FROM N.A.
 RC TISSUE=Skin;
 RX MEDLINE=20062965; PubMed=10594734;
 RA Whitcock N.V., Ashton G.H., Dopping-Hepenstal P.J., Gratian M.J.,
 RA Keane F.M., Eady R.A.J., McGrath J.A.;
 RT "Striate palmoplantar keratoderma resulting from desmoplakin
 haploinsufficiency.";
 RL J. Invest. Dermatol. 113:940-946(1999).
 RN [6]
 RP CHARACTERIZATION.
 RX MEDLINE=98012209; PubMed=9348293;
 RA Kowalczyk A.P., Bornslaeger E.A., Borgwardt J.E., Palka H.L.,
 RA Dhalwal A.S., Corcoran C.M., Denning M.F., Green K.J.;
 RT "The amino-terminal domain of desmoplakin binds to plakoglobin and
 clusters desmosomal cadherin-plakoglobin complexes.";
 RL J. Cell Biol. 139:773-784(1997).
 RN [7]
 RP LIPIDATION.
 RX MEDLINE=98316349; PubMed=9651377;
 RA Marekov L.N., Steinert P.M.;
 RT "Ceramides are bound to structural proteins of the human foreskin
 epidermal cornified cell envelope.";
 RL J. Biol. Chem. 273:17763-17770(1998).
 RN [8]
 RP VARIANT ARVD8 ARG-299.
 RX MEDLINE=22285852; PubMed=12373648;
 RA Rampazzo A., Nava A., Malacrida S., Beggagna G., Baucé B., Rossi V.,
 RA Zimbello R., Simionati B., Basso C., Thiene G., Towbin J.A.,
 RA Danieli G.A.;
 RT "Mutation in human desmoplakin domain binding to plakoglobin causes a
 dominant form of arrhythmogenic right ventricular cardiomyopathy.";
 RL Am. J. Hum. Genet. 71:1200-1206(2002).
 RN [9]
 RP VARIANTS SFWS LYS-287 AND CYS-2366.
 RX MEDLINE=21830938; PubMed=11841536;
 RA Whitcock N.V., Wan H., Morley S.M., Garzon M.C., Kristal L., Hyde P.,
 RA McLean W.H.I., Pulkkinen L., Uitto J., Cristiano A.M., Eady R.A.J.,
 RA McGrath J.A.;
 RT "Compound heterozygosity for non-sense and mis-sense mutations in
 desmoplakin underlies skin fragility/woolly hair syndrome.";
 RL J. Invest. Dermatol. 118:232-238(2002).
 CC -!- FUNCTION: Major high molecular weight protein of desmosomes.
 CC Involved in the organization of the desmosomal cadherin-
 CC plakoglobin complexes into discrete plasma membrane domains and in
 CC the anchoring of intermediate filaments to the desmosomes.
 CC -!- SUBUNIT: Homodimer.
 CC -!- SUBCELLULAR LOCATION: Innermost portion of the desmosomal plaque.
 CC -!- ALTERNATIVE PRODUCTS:
 CC Event-Alternative splicing; Named isoforms=2;
 CC Name=DPI; Synonyms=Dp1;
 CC IsoId=P15924-1, Sequence=Displayed;
 CC Name=DPII; Synonyms=DP2;
 CC IsoId=P15924-2; Sequence=VSP_005070;
 CC -!- TISSUE SPECIFICITY: Isoform DPI is apparently an obligate
 CC constituent of all desmosomes; Isoform DPII resides predominantly
 CC in tissues and cells of stratified origin.
 CC -!- DOMAIN: The N-terminal region is required for localization to the
 CC desmosomal plaque and interacts with the N-terminal region of
 CC plakophilin 1. The C-terminal region interacts with intermediate
 CC filaments.
 CC -!- PTM: Substrate of transglutaminase. Some glutamines and lysines
 CC are cross-linked to other desmoplakin molecules, to other proteins
 CC such as keratin, envoplakin, periplakin and involucrin, and to
 CC lipids like omega-hydroxyceramide.
 CC -!- DISEASE: Defects in DSP are the cause of familial arrhythmogenic
 CC right ventricular dysplasia-8 (ARVD8) [MIM:607450], an autosomal
 CC dominant disorder.
 CC -!- DISEASE: Defects in DSP are a cause of striate palmoplantar
 CC keratoderma II (KPPK2, KPPS2 or SPPK2), characterized by skin
 CC thickening in the palms (linear pattern) and the soles (island-
 CC like pattern) and flexor aspect of the fingers; and rarely by
 CC abnormalities of the nails, the teeth and the hair.
 CC -!- DISEASE: Defects in DSP are the cause of skin fragility-woolly
 CC hair syndrome (SFWS) [MIM:607655], an autosomal recessive
 CC genodermatosis characterized by focal and diffuse palmoplantar
 CC keratoderma, hyperkeratotic plaques on the trunk and limbs, and
 CC woolly hair with varying degrees of alopecia.
 CC -!- SIMILARITY: Contains 17 plectrin repeats.
 CC -!- SIMILARITY: Belongs to the plakin or cytolinker family.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL; M77830; AAA85135.1; -;
 CC EMBL; AL031058; -; NOT ANNOTATED_CDS.
 CC EMBL; J05211; AAA35766.1; -;
 CC EMBL; AF139065; AAF19785.1; -;
 CC PIR; A38194; A38194.
 CC PDB; 1LM5; 31-JUL-02.
 CC PDB; 1LM7; 31-JUL-02.
 CC Genew; HGNC:3052; DSP.
 CC MIM; 125647; -;
 CC MIM; 607450; -;
 CC MIM; 607655; -;
 CC GO; GO:0005200; P:structural constituent of cytoskeleton; TAS.
 CC GO; GO:0008444; P:epidermal differentiation; TAS.
 CC InterPro; IPR001101; Plectrin repeat.
 CC InterPro; IPR002017; Spectrin.
 CC Pfam; PF00681; Plectrin; 11.


```

Query Match          9.3%; Score 252.5; DB 1; Length 880;
Best Local Similarity 23.2%; Pred. No. 0.0017;
Matches 123; Conservative 98; Mismatches 183; Indels 127; Gaps 22;

QY 12 EGAQRPSQAAPVAEAGPGSSQ-----APRKPEGAQACATAGGALRDVSHLSQLE 64
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 259 KGLEKIVQIERSIERKAKISELSEIVKDIPIKQEKKEVYKRLKG-PRDEYESKRLLE 317
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 65 DILSTYCVUNNGGFGDGAQGEAPEDAKSRITYVARNGEPEPTPVVYGEKPSKGP 124
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 318 KELSWM-----ESLKAEIEVKEGE-----KKKE 342
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 125 NTEIRQS-DEYGR-DHRRP--QEKKAAGLGEITLLMOTLNTLSTPE--EKLAALCK 178
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 343 RAEIRKLESEKLEELKPYVELEDAKQVQKQIBLKXKLSLSPGEVIEKLESLEK 402
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 179 KYAELEERHNSOKWMLQKQSOLOVQKHRLGHSKAVLAKLSLIC-----REL-Q 233
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 403 ERTET-----BEAKITTRIGOMQEK-----ERMKAEILKAKGKPCVCGRELTE 451
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 234 RHNRSLK-----BEGVQAREEKEKEVTSHPQVTLNDIQLQ--MEQHNRNS 280
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 452 EKKELMERYTLEIKKIBELKRTTEERKLAVNLKLEIKLRPSVWEDIAEQIKLES 511
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 281 KLRQENELAE-----RLKLIQVELBEHIDIKVFK 312
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 512 KLKGNLELEQKEREFEGLANBPNKGGELLGLERDLKRIKALEGRKRLTEKVRKAKE 571
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 313 H-KDQQQLVDKLAQQAQM---LKEAEERHOREKDFLLKAEVSRQMSCLMKQOETHLK 368
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 572 ELENLHRLQRLGPGFSEVLELNLRIGLEEFHDKYVKAKESS-ELRELKQKLEKEKTELD 630
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 369 QQALYTERKFEFPQNTLSKSEVFTTP-KQEMRWTKIKKLEKETTMYRSWESSNKA- 426
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 631 QAFEMLAUVENIEFKAKLDLESKFNBYEYKREKRLVLEBSVSLARLEELKSKV 690
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 427 -----LLEMAEKTVRDKLEGLQVQIKQRLKCLRALQTRNDLNRKVPD 471
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 691 EQIKATRLKLEKEEREK-----AKLEIKLEKALKSKE-----DLRKKIRD 733
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :

RESULT 8
TRHY SHEEP
ID TRHY SHEEP STANDARD; PRT; 1549 AA.
AC
DT 01-AUG-1991 (Rel. 19, Created)
DT 01-OCT-1994 (Rel. 30, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Trichohyalin.
GN THH.
OS Ovis aries (Sheep).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
OC Bovidae; Caprinae; Ovis.
OX NCBI_TaxID=9940;
RN [1]_TaxID=9940;
RP SEQUENCE FROM N.A.
RX MEDLINE=93260018; PubMed=7684041;
RA Fietz M.J., McLaughlin C.J., Campbell M.T., Rogers G.E.;
RT "Analysis of the sheep trichohyalin gene: potential structural and
RT calcium-binding roles of trichohyalin in the hair follicle.";
RL J. Cell Biol. 121:855-865(1993).
RN [2]
RP SEQUENCE OF 1016-1549 FROM N.A.
RC STRAIN=Merino-Dorset horn X Border Leicester; TISSUE=Wool follicles;
RX MEDLINE=90130632; PubMed=2298812;
RA Fietz M.J., Presland R.B., Rogers G.E.;
RT "The cDNA-deduced amino acid sequence for trichohyalin, a
RT differentiation marker in the hair follicle, contains a 23 amino acid
RT repeat.";
RL J. Cell Biol. 110:427-436(1990).
CC -!- FUNCTION: Intermediate filament-associated protein that associates

```

```

in regular arrays with keratin intermediate filaments (KIF) of the
inner root sheath cells of the hair follicle and the granular
layer of the epidermis. It later becomes cross-linked to KIF by
isodisulphide bonds. It may serve as scaffold protein, together
with involucrin, in the organization of the cell envelope or even
anchor the cell envelope to the KIF network. It may be involved in
its own calcium-dependent postsynthetic processing during terminal
differentiation.
-!- SUBUNIT: Homodimer (Probable).
-!- ALTERNATIVE PRODUCTS:
Event=Alternative splicing; Named isoforms=2;
Comment=Additional isoforms seem to exist;
Name=Long;
IsoID=P22793-1; Sequence=Displayed;
Name=Short;
IsoID=P22793-2; Sequence=VSP 000847, VSP 000848;
-!- TISSUE SPECIFICITY: Found in the hard keratinizing tissues such as
the inner root sheath (IRS) of hair follicles and medulla, and in
the epithelia of the tongue, hoof and rumen.
-!- DOMAIN: Consists of nine domains. Domain 1 contains two EF-hand
calcium-binding domains. Domains 2-4, 6, and 8 are almost entirely
alpha-helical, configured as a series of peptide repeats of
varying regularity, and are thought to form a single-stranded
alpha-helical rod stabilized by ionic interactions. Domain 6 is
the most regular and may bind KIF directly by ionic interactions.
Domains 5 and 7 are less well organized and may induce folds in
the molecule. Domain 9 contains the C-terminus, conserved among
different species.
-!- PTM: Substrate of transglutaminase. Some 200 arginines are
probably converted to citrullines by peptidylarginine deiminase.
-!- SIMILARITY: In the N-terminal section; belongs to the S-100
family.
-!- SIMILARITY: Contains 2 EF-hand calcium-binding domains.
This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by not-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
EMBL; Z18361; CAA79165.1; -
EMBL; X51695; CAA35992.1; -
PIR; A40691; A40691.
DR HSSP; P02633; 1IG5.
InterPro; IPR001751; CAPS S100.
InterPro; IPR002048; EF-hand.
Pfam; PF00036; ehand; 1.
Pfam; PF01023; S_100; 1.
ProDom; PD003407; CAPS S100; 1.
PROSITE; PS00018; EF HAND; 1.
PROSITE; PS00303; S100 CAPS; FALSE NEG.
KW Keratinization; Calcium-binding; Repeat; Citrullination;
KW Alternative splicing.
FT DOMAIN 1 91 S-100 LIKE
FT CA_BIND 22 33 EF-HAND 1 (LOW AFFINITY) (POTENTIAL).
FT CA_BIND 62 73 EF-HAND 2 (HIGH AFFINITY) (POTENTIAL).
FT DOMAIN 413 832 14 X 28 AA APPROXIMATE TANDEM REPEATS.
FT REPEAT 413 448 1-1.
FT REPEAT 449 476 1-2.
FT REPEAT 477 504 1-3.
FT REPEAT 505 532 1-4.
FT REPEAT 533 560 1-5.
FT REPEAT 561 588 1-6.
FT REPEAT 589 616 1-7.
FT REPEAT 617 644 1-8.
FT REPEAT 645 678 1-9.
FT REPEAT 679 706 1-10.
FT REPEAT 707 742 1-11.
FT REPEAT 743 771 1-12.
FT REPEAT 772 796 1-13.
FT REPEAT 797 832 1-14.

```



```

DOMAIN          938 1507
REPEAT          938 961
REPEAT          962 985
REPEAT          986 1021
REPEAT          1022 1044
REPEAT          1045 1067
REPEAT          1068 1090
REPEAT          1091 1121
REPEAT          1122 1144
REPEAT          1145 1167
REPEAT          1168 1197
REPEAT          1198 1227
REPEAT          1228 1250
REPEAT          1251 1273
REPEAT          1274 1296
REPEAT          1297 1319
REPEAT          1320 1342
REPEAT          1343 1368
REPEAT          1369 1391
REPEAT          1392 1416
REPEAT          1417 1439
REPEAT          1440 1461
REPEAT          1462 1484
REPEAT          1485 1507
REPEAT          1508 1529
REPEAT          1530 1551
REPEAT          1552 1573
REPEAT          1574 1595
REPEAT          1596 1617
REPEAT          1618 1639
REPEAT          1640 1661
REPEAT          1662 1683
REPEAT          1684 1705
REPEAT          1706 1727
REPEAT          1728 1749
REPEAT          1750 1771
REPEAT          1772 1793
REPEAT          1794 1815
REPEAT          1816 1837
REPEAT          1838 1859
REPEAT          1860 1881
REPEAT          1882 1903
REPEAT          1904 1925
REPEAT          1926 1947
REPEAT          1948 1969
REPEAT          1970 1991
REPEAT          1992 2012
REPEAT          2013 2034
REPEAT          2035 2055
REPEAT          2056 2076
REPEAT          2077 2098
REPEAT          2099 2120
REPEAT          2121 2141
REPEAT          2142 2162
REPEAT          2163 2183
REPEAT          2184 2204
REPEAT          2205 2225
REPEAT          2226 2246
REPEAT          2247 2267
REPEAT          2268 2288
REPEAT          2289 2309
REPEAT          2310 2330
REPEAT          2331 2351
REPEAT          2352 2372
REPEAT          2373 2393
REPEAT          2394 2414
REPEAT          2415 2435
REPEAT          2436 2456
REPEAT          2457 2477
REPEAT          2478 2498
REPEAT          2499 2519
REPEAT          2520 2540
REPEAT          2541 2561
REPEAT          2562 2582
REPEAT          2583 2603
REPEAT          2604 2624
REPEAT          2625 2645
REPEAT          2646 2666
REPEAT          2667 2687
REPEAT          2688 2708
REPEAT          2709 2729
REPEAT          2730 2750
REPEAT          2751 2771
REPEAT          2772 2792
REPEAT          2793 2813
REPEAT          2814 2834
REPEAT          2835 2855
REPEAT          2856 2876
REPEAT          2877 2897
REPEAT          2898 2918
REPEAT          2919 2939
REPEAT          2940 2960
REPEAT          2961 2981
REPEAT          2982 3002
REPEAT          3003 3023
REPEAT          3024 3044
REPEAT          3045 3065
REPEAT          3066 3086
REPEAT          3087 3107
REPEAT          3108 3128
REPEAT          3129 3149
REPEAT          3150 3170
REPEAT          3171 3191
REPEAT          3192 3212
REPEAT          3213 3233
REPEAT          3234 3254
REPEAT          3255 3275
REPEAT          3276 3296
REPEAT          3297 3317
REPEAT          3318 3338
REPEAT          3339 3359
REPEAT          3360 3380
REPEAT          3381 3401
REPEAT          3402 3422
REPEAT          3423 3443
REPEAT          3444 3464
REPEAT          3465 3485
REPEAT          3486 3506
REPEAT          3507 3527
REPEAT          3528 3548
REPEAT          3549 3569
REPEAT          3570 3590
REPEAT          3591 3611
REPEAT          3612 3632
REPEAT          3633 3653
REPEAT          3654 3674
REPEAT          3675 3695
REPEAT          3696 3716
REPEAT          3717 3737
REPEAT          3738 3758
REPEAT          3759 3779
REPEAT          3780 3800
REPEAT          3801 3821
REPEAT          3822 3842
REPEAT          3843 3863
REPEAT          3864 3884
REPEAT          3885 3905
REPEAT          3906 3926
REPEAT          3927 3947
REPEAT          3948 3968
REPEAT          3969 3989
REPEAT          3990 4010
REPEAT          4011 4031
REPEAT          4032 4052
REPEAT          4053 4073
REPEAT          4074 4094
REPEAT          4095 4115
REPEAT          4116 4136
REPEAT          4137 4157
REPEAT          4158 4178
REPEAT          4179 4199
REPEAT          4200 4220
REPEAT          4221 4241
REPEAT          4242 4262
REPEAT          4263 4283
REPEAT          4284 4304
REPEAT          4305 4325
REPEAT          4326 4346
REPEAT          4347 4367
REPEAT          4368 4388
REPEAT          4389 4409
REPEAT          4410 4430
REPEAT          4431 4451
REPEAT          4452 4472
REPEAT          4473 4493
REPEAT          4494 4514
REPEAT          4515 4535
REPEAT          4536 4556
REPEAT          4557 4577
REPEAT          4578 4598
REPEAT          4599 4619
REPEAT          4620 4640
REPEAT          4641 4661
REPEAT          4662 4682
REPEAT          4683 4703
REPEAT          4704 4724
REPEAT          4725 4745
REPEAT          4746 4766
REPEAT          4767 4787
REPEAT          4788 4808
REPEAT          4809 4829
REPEAT          4830 4850
REPEAT          4851 4871
REPEAT          4872 4892
REPEAT          4893 4913
REPEAT          4914 4934
REPEAT          4935 4955
REPEAT          4956 4976
REPEAT          4977 4997
REPEAT          4998 5018
REPEAT          5019 5039
REPEAT          5040 5060
REPEAT          5061 5081
REPEAT          5082 5102
REPEAT          5103 5123
REPEAT          5124 5144
REPEAT          5145 5165
REPEAT          5166 5186
REPEAT          5187 5207
REPEAT          5208 5228
REPEAT          5229 5249
REPEAT          5250 5270
REPEAT          5271 5291
REPEAT          5292 5312
REPEAT          5313 5333
REPEAT          5334 5354
REPEAT          5355 5375
REPEAT          5376 5396
REPEAT          5397 5417
REPEAT          5418 5438
REPEAT          5439 5459
REPEAT          5460 5480
REPEAT          5481 5501
REPEAT          5502 5522
REPEAT          5523 5543
REPEAT          5544 5564
REPEAT          5565 5585
REPEAT          5586 5606
REPEAT          5607 5627
REPEAT          5628 5648
REPEAT          5649 5669
REPEAT          5670 5690
REPEAT          5691 5711
REPEAT          5712 5732
REPEAT          5733 5753
REPEAT          5754 5774
REPEAT          5775 5795
REPEAT          5796 5816
REPEAT          5817 5837
REPEAT          5838 5858
REPEAT          5859 5879
REPEAT          5880 5900
REPEAT          5901 5921
REPEAT          5922 5942
REPEAT          5943 5963
REPEAT          5964 5984
REPEAT          5985 6005
REPEAT          6006 6026
REPEAT          6027 6047
REPEAT          6048 6068
REPEAT          6069 6089
REPEAT          6090 6110
REPEAT          6111 6131
REPEAT          6132 6152
REPEAT          6153 6173
REPEAT          6174 6194
REPEAT          6195 6215
REPEAT          6216 6236
REPEAT          6237 6257
REPEAT          6258 6278
REPEAT          6279 6299
REPEAT          6300 6320
REPEAT          6321 6341
REPEAT          6342 6362
REPEAT          6363 6383
REPEAT          6384 6404
REPEAT          6405 6425
REPEAT          6426 6446
REPEAT          6447 6467
REPEAT          6468 6488
REPEAT          6489 6509
REPEAT          6510 6530
REPEAT          6531 6551
REPEAT          6552 6572
REPEAT          6573 6593
REPEAT          6594 6614
REPEAT          6615 6635
REPEAT          6636 6656
REPEAT          6657 6677
REPEAT          6678 6698
REPEAT          6699 6719
REPEAT          6720 6740
REPEAT          6741 6761
REPEAT          6762 6782
REPEAT          6783 6803
REPEAT          6804 6824
REPEAT          6825 6845
REPEAT          6846 6866
REPEAT          6867 6887
REPEAT          6888 6908
REPEAT          6909 6929
REPEAT          6930 6950
REPEAT          6951 6971
REPEAT          6972 6992
REPEAT          6993 7013
REPEAT          7014 7034
REPEAT          7035 7055
REPEAT          7056 7076
REPEAT          7077 7097
REPEAT          7098 7118
REPEAT          7119 7139
REPEAT          7140 7160
REPEAT          7161 7181
REPEAT          7182 7202
REPEAT          7203 7223
REPEAT          7224 7244
REPEAT          7245 7265
REPEAT          7266 7286
REPEAT          7287 7307
REPEAT          7308 7328
REPEAT          7329 7349
REPEAT          7350 7370
REPEAT          7371 7391
REPEAT          7392 7412
REPEAT          7413 7433
REPEAT          7434 7454
REPEAT          7455 7475
REPEAT          7476 7496
REPEAT          7497 7517
REPEAT          7518 7538
REPEAT          7539 7559
REPEAT          7560 7580
REPEAT          7581 7601
REPEAT          7602 7622
REPEAT          7623 7643
REPEAT          7644 7664
REPEAT          7665 7685
REPEAT          7686 7706
REPEAT          7707 7727
REPEAT          7728 7748
REPEAT          7749 7769
REPEAT          7770 7790
REPEAT          7791 7811
REPEAT          7812 7832
REPEAT          7833 7853
REPEAT          7854 7874
REPEAT          7875 7895
REPEAT          7896 7916
REPEAT          7917 7937
REPEAT          7938 7958
REPEAT          7959 7979
REPEAT          7980 8000
REPEAT          8001 8021
REPEAT          8022 8042
REPEAT          8043 8063
REPEAT          8064 8084
REPEAT          8085 8105
REPEAT          8106 8126
REPEAT          8127 8147
REPEAT          8148 8168
REPEAT          8169 8189
REPEAT          8190 8210
REPEAT          8211 8231
REPEAT          8232 8252
REPEAT          8253 8273
REPEAT          8274 8294
REPEAT          8295 8315
REPEAT          8316 8336
REPEAT          8337 8357
REPEAT          8358 8378
REPEAT          8379 8399
REPEAT          8400 8420
REPEAT          8421 8441
REPEAT          8442 8462
REPEAT          8463 8483
REPEAT          8484 8504
REPEAT          8505 8525
REPEAT          8526 8546
REPEAT          8547 8567
REPEAT          8568 8588
REPEAT          8589 8609
REPEAT          8610 8630
REPEAT          8631 8651
REPEAT          8652 8672
REPEAT          8673 8693
REPEAT          8694 8714
REPEAT          8715 8735
REPEAT          8736 8756
REPEAT          8757 8777
REPEAT          8778 8798
REPEAT          8799 8819
REPEAT          8820 8840
REPEAT          8841 8861
REPEAT          8862 8882
REPEAT          8883 8903
REPEAT          8904 8924
REPEAT          8925 8945
REPEAT          8946 8966
REPEAT          8967 8987
REPEAT          8988 9008
REPEAT          9009 9029
REPEAT          9030 9050
REPEAT          9051 9071
REPEAT          9072 9092
REPEAT          9093 9113
REPEAT          9114 9134
REPEAT          9135 9155
REPEAT          9156 9176
REPEAT          9177 9197
REPEAT          9198 9218
REPEAT          9219 9239
REPEAT          9240 9260
REPEAT          9261 9281
REPEAT          9282 9302
REPEAT          9303 9323
REPEAT          9324 9344
REPEAT          9345 9365
REPEAT          9366 9386
REPEAT          9387 9407
REPEAT          9408 9428
REPEAT          9429 9449
REPEAT          9450 9470
REPEAT          9471 9491
REPEAT          9492 9512
REPEAT          9513 9533
REPEAT          9534 9554
REPEAT          9555 9575
REPEAT          9576 9596
REPEAT          9597 9617
REPEAT          9618 9638
REPEAT          9639 9659
REPEAT          9660 9680
REPEAT          9681 9701
REPEAT          9702 9722
REPEAT          9723 9743
REPEAT          9744 9764
REPEAT          9765 9785
REPEAT          9786 9806
REPEAT          9807 9827
REPEAT          9828 9848
REPEAT          9849 9869
REPEAT          9870 9890
REPEAT          9891 9911
REPEAT          9912 9932
REPEAT          9933 9953
REPEAT          9954 9974
REPEAT          9975 9995
REPEAT          9996 10000

```

Query Match 9.3%; Score 251.5; DB 1; Length 1549;

Best Local Similarity 22.7%; Pred. No. 0.0032;

Matches 111; Conservative 108; Mismatches 194; Indels 77; Gaps 17;

6 QPEAGPGAGRPSQAAPAYEAGPGSSQAPRPECAQAATAGSGALRDVSELSQGLD 65

521 QREERKREERERQYKLEVEQLQREERKREERKQ-----YLRKLEVE 574

66 ILSTYCVNNQGGPDEAGAGPABEDAKSTYVARNGEPPTVPVYGEKPSKDPN 125

575 QL-----QRERKREERERQYKLEVEQLQREERKREERKQ 617

126 TE-----EIQSDVGDHRRPQKKAGLKGKETTLLMOTLNTLSTPEKLAALCKY 180

618 RYKLEVEQLQREERQYKLEVEQLQREERKREERKQ-----LQREERL 663

181 AELLERNSQKMLQKQKQKQKQKQKQKQKQKQKQKQKQKQKQKQKQKQKQKQ 239

664 EQLLEERERKREERQYKLEVEQLQREERKREERKQKQKQKQKQKQKQKQKQ 723

240 KEQGVQARERERKREERKREERKREERKREERKREERKREERKREERKREER 298

724 QREKQQLQREERKREERKREERKREERKREERKREERKREERKREERKREER 780

299 QYELREBEHDKVFKHDLQQLQKQKQKQKQKQKQKQKQKQKQKQKQKQKQKQ 358

781 QRE--EQLQREERKREERKQ-----YREERKREERKREERKQ--LQREERK 831

359 LMQQETHLQKQALYKTEPEP-----QNTLSKSEVFTTFKQEMKWKIKKLEK 412

832 LEQLBEELQRL-----DKRQFRDDQKQNEV-RNSRYVSKHREKKS-----KQD-D 881

413 TTYRGRWSSNKALLEABETVRDKLEGLQVK-----IQRLKCALQGTERRDLAK 467

882 SWRESQFQDLPLQDEQEKREERKREERKREERKREERKREERKREERKREER 941

468 RVQDLSAGGQ 477

942 REEQILLKQ 951

RA Harris N.L., Richter J., Russo S., Schroeder A.J., Shu S.Q.,
RA Stapleton M., Yamada C., Ashburner M., Gelbart W.M., Rubin G.M.,
RA Lewis S.E.,
RT "Annotation of the Drosophila melanogaster euchromatic genome: a
RT systematic review.",
RL Genome Biol. 3:RESEARCH0083.1-RESEARCH0083.22(2002).
RN [4]
RP SEQUENCE FROM N.A. (ISOFORM B).
RC STRAIN=Berkeley; TISSUE=Embryo;
RX MEDLINE=22426066; PubMed=12537569;
RA Stapleton M., Carlson J.W., Brokstein P., Yu C., Champe M.,
RA George R.A., Guerin H., Krommiller B., Pacleb J.M., Park S., Wan K.H.,
RA Rubin G.M., Celniker S.E.,
RT "A Drosophila full-length cDNA resource.",
RL Genome Biol. 3:RESEARCH0080.1-RESEARCH0080.8(2002).
RN [5]
RP FUNCTION, SUBUNIT, SUBCELLULAR LOCATION, AND TISSUE SPECIFICITY.
RX MEDLINE=20530668; PubMed=11076973;
RA Sisson J.C., Field C., Ventura R., Royou A., Sullivan W.,
RT "Lava lamp, a novel peripheral Golgi protein, is required for
RT Drosophila melanogaster cellularization.",
RL J. Cell Biol. 151:905-918(2000).
CC -!- FUNCTION: Together CLIP-190 and jar may coordinate the interaction
CC between the actin and microtubule cytoskeleton. May link endocytic
CC vesicles to microtubules. May play a role in formation of furrows
CC during cellularization.
CC -!- SUBUNIT: Interacts with Lva.
CC -!- SUBCELLULAR LOCATION: Cytoplasmic; microtubule-associated. Lva-
CC CLIP-190 complexes are found at the Golgi.
CC -!- ALTERNATIVE PRODUCTS:
CC Event=Alternative splicing; Named isoforms=3;
CC Name=A;
CC IsoId=Q9VJE5-1; Sequence=Displayed;
CC Name=B;
CC IsoId=Q9VJE5-2; Sequence=VSP_050479;
CC Note=No experimental confirmation available;
CC Name=C;
CC IsoId=Q9VJE5-3; Sequence=VSP_050480;
CC Note=No experimental confirmation available;
CC -!- TISSUE SPECIFICITY: Specifically expressed at the tip of the
CC furrow in cellularizing blastoderms. CLIP-190 and jar are
CC coexpressed at several times in development and in a number of
CC tissues, including embryonic axonal neuron processes and posterior
CC pole.
CC -!- SIMILARITY: Contains 2 CAP-Gly domains.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; AF041182; AAB96783.1; -
CC EMBL; AE003655; AAF53604.1; -
CC EMBL; AE003655; AAF53605.2; -
CC EMBL; AE003655; AAN10987.1; -
CC EMBL; AY118896; AAM50756.1; -
CC FlyBase; FBGN020503; CLIP-190.
CC GO; GO:0005938; C:cell cortex; IDA.
CC GO; GO:0005794; C:Golgi apparatus; IDA.
CC GO; GO:0005875; C:microtubule associated complex; IDA.
CC GO; GO:0003779; F:actin binding; IDA.
CC GO; GO:0008017; F:microtubule binding; IDA.
CC GO; GO:0005515; F:protein binding; IPI.
CC GO; GO:0007349; P:cellularization; IMP.
CC InterPro; IPR000938; CAP-Gly.
CC Pfam; PF01302; CAP_Gly; 2.
CC PROSITE; PS00845; CAP_Gly_1; FALSE_NEG.
CC PROSITE; PS0245; CAP_Gly_2; 2.
KW Cytoskeleton; Golgi stack; Microtubule; Actin-binding; Coiled coil;
KW Repeat; Alternative splicing.

FT	DOMAIN	143	185	CAP-GLY 1.
FT	DOMAIN	260	302	CAP-GLY 2.
FT	DOMAIN	378	468	COILED COIL (POTENTIAL).
FT	DOMAIN	484	660	COILED COIL (POTENTIAL).
FT	DOMAIN	488	1452	GLV-RICH.
FT	DOMAIN	667	916	COILED COIL (POTENTIAL).
FT	DOMAIN	926	981	COILED COIL (POTENTIAL).
FT	DOMAIN	1001	1121	COILED COIL (POTENTIAL).
FT	DOMAIN	1158	1549	COILED COIL (POTENTIAL).
FT	DOMAIN	1200	1461	GLN-RICH.
FT	DOMAIN	1565	1600	COILED COIL (POTENTIAL).
FT	VARSPIC	348		Missing (in isoform B).
FT	VARSPIC			/FTId=VSP_050479.
FT				MDDTSASGTSAPSPSPVTADPEPCATSKLPGPIRSNIP
FT				TEATSGTGIPQSKMKAPSPSGTSGVSKIGPCNCHTTPK
FT				SGPPPEATSMRESDNLSINSAT -> MSRESDNLS
FT				SINSAITDLYQTVRRFTSSLSPTDMDRFPSPARSLKSE
FT				AGRSADYLLATGRRSS (in isoform
FT				C).
FT				/FTId=VSP_050480.
FT				S -> N (IN REF. 1).
FT				D -> G (IN REF. 1).
FT				K -> Q (IN REF. 1).
FT				E -> A (IN REF. 1).
FT				T -> S (IN REF. 1).
FT				S -> I (IN REF. 1).
FT				N -> Q (IN REF. 1).
FT				M -> K (IN REF. 1).
FT				F -> L (IN REF. 1).
FT				Q -> E (IN REF. 1).
FT				HLL -> QLQ (IN REF. 1).
FT				Q -> E (IN REF. 1).
FT				G -> E (IN REF. 1).
FT				C -> Y (IN REF. 1).
SQ	SEQUENCE	1690 AA; 189063 MW; D5F7916A9C532F16 CRC64;		
	Query Match	9.3%; Score 251.5; DB 1; Length 1690;		
	Best Local Similarity	23.3%; Pred. No. 0.0035;		
	Matches 127; Conservative	102; Mismatches 180; Indels 135; Gaps 23;		
QY	33	SOAPRPEGAQARTAQSGALRDVBSBELSQLEDILSTYCDVNNQGGFGDGAQGEPAEPE	92	
DB	352	SSTPVPILATPKSQFS--MQDLLREKQHQVEKLAVERDLDR-----E	392	
QY	93	DAKSRITYVARNGEPEPTPVYGEKEPKSGDPNTEIROS-DEV---GDNRHRPO-EKK	147	
DB	393	DAQNOALQKQNTINELKARIVELSEALDNERKKTLEQLCSIDEAQFCGDELNAQSQVYKE	452	
QY	148	KAGLCKEITLMQ---TINTLSTP-----EKLALCKK-----VAELL	184	
DB	453	KIHDLESKITKLVSATPSLQSIPLPDDGALQEEIAKLQKQMTIQQEVESRIAEQL	512	
QY	185	BEHRNSQKQKLLQKK---QSOLVQEKDHLRGEHSAVLARSKLESCLRELQHRNSLK	240	
DB	513	EEBQRLRENVKYLNEQIATLQSELVSKDEAL---EKFSLSBEGCIENLRLEL---	564	
QY	241	EEGVQAPAREE-----EKREY-----TSHFQVTLNDI	268	
DB	565	EENEQAQEAQAEFTFKLAESVEVLRLSELQNLKATSDLESERVNKTDECEILQTEV	624	
QY	269	QLQWEQHNRNSKL---RQENMELAE-----RLKLLIROYELREHIDKVFVKHOL	316	
DB	625	RMEDEQIRELNQQLDEVITQLNVQKADSSALDDMLRLQK--EGTEEKSTLLEKT----	676	
QY	317	QOQLVDAKIQQAQOEMLEKEBERHOREKDFLLKEAVESQRMCELMQO-----QETHLK	368	
DB	677	EKELVQSK-EQAATKLNDEKLEKQISD--LKQAEQEKLVREMTENAINQIOLEKESIE	733	
QY	369	QQLALVTEKFEFQNTLSKSSSEVFTPFQO-----EMEKMTYKIKKLEKETTMYRSWE	421	
DB	734	QQLALKQNELEDPKQKQSEVHLEIKAQNTQKDFELVESGSLKKLQQLQEQKTLGHE	793	
QY	422	SSNKALLEMAEKTIV-----RDKLEGLQVKIQRLEKLCRALQTERNDLNKRVQDLSAGGQ	477	

Db 794 KQALSELKKEKTIKKEQELQLOSKAESALKVQVLEQLQO--QAAAGEE 851
 QY 478 GSLT 481
 Db 852 GSKT 855

RESULT 10

MYH4 RABIT
 ID MYH4 RABIT STANDARD; PRT; 1938 AA.
 AC Q28641;
 DT 16-OCT-2001 (Rel. 40, Created)
 DT 16-OCT-2001 (Rel. 40, Last sequence update)
 DT 15-MAR-2004 (Rel. 43, Last annotation update)
 DE Myosin heavy chain, skeletal muscle, juvenile.
 OS Oryctolagus cuniculus (Rabbit).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Lagomorpha; Leporidae; Oryctolagus.
 OX NCBI_TaxID=9986;
 RN [1]
 PP SEQUENCE FROM N.A.
 RC STRAIN=New Zealand white; TISSUE=Skeletal muscle;
 RA Maeda K., Hostinova E., Roesch-Klein Kauf A., Schuster H., Gasperik J.,
 RA Wittinchofer A.;
 RT skeletal muscle and a novel cosynthesis of S-1 fragment with the
 RT essential and regulatory light chains.";
 RL Submitted (AUG-1995) to the EMBL/GenBank/DBJ databases.
 CC -!- FUNCTION: Muscle contraction.
 CC -!- SUBUNIT: Muscle myosin is a hexameric protein that consists of 2
 CC heavy chain subunits (MHC), 2 alkali light chain subunits (MLC)
 CC and 2 regulatory light chain subunits (MLC-2).
 CC -!- SUBCELLULAR LOCATION: Thick filaments of the myofibrils.
 CC -!- DOMAIN: The rodlike tail sequence is highly repetitive, showing
 CC cycles of a 28-residue repeat pattern composed of 4 heptapeptides,
 CC characteristic for alpha-helical coiled coils.
 CC -!- MISCELLANEOUS: Each myosin heavy chain can be split into 1 light
 CC meromyosin (LM) and 1 heavy meromyosin (HMM). It can later be
 CC split further into 2 globular subfragments (S1) and 1 rod-shaped
 CC subfragment (S2).
 CC -!- SIMILARITY: Contains 1 myosin-like globular head domain.
 CC -!- SIMILARITY: Contains 1 IQ domain.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation
 CC at the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).

EMBL; U32574; AAA74199.1; -;
 DR F1R; A59293; A59293.
 DR HSSP; P13538; 2MYS.
 DR InterPro; IPR000048; IQ region.
 DR InterPro; IPR001609; myosin head.
 DR InterPro; IPR004009; Myosin_N.
 DR InterPro; IPR002928; Myosin_tail.
 DR Pfam; PF00612; IQ; 2.
 DR Pfam; PF00063; myosin head; 1.
 DR Pfam; PF02736; Myosin_N; 1.
 DR Pfam; PF01576; Myosin_tail; 1.
 DR PRINTS; PR00193; MYOSINHEAVY.
 DR ProDom; PD000355; myosin_head; 1.
 DR SMART; SM00015; IQ; 1.
 DR SMART; SM00242; MYSC; 1.
 DR PROSITE; PS00096; IQ; 1.
 DR Myosin; Muscle protein; Coiled coil; Thick filament; Actin-binding;
 KW Calmodulin-binding; ATP-binding; Methylation; Multigene family.
 DOMAIN 1 783
 FT MYOSIN HEAD-LIKE.
 FT 784 813
 FT 842 1938
 FT COILED COIL (POTENTIAL).

FT NP BIND 179 186 ATP (POTENTIAL).
 FT DOMAIN 658 680 ACTIN-BINDING (BY SIMILARITY).
 FT DOMAIN 760 774 ACTIN-BINDING (BY SIMILARITY).
 FT MOD RES 35 35 METHYLATION (MONO-) (BY SIMILARITY).
 FT MOD RES 130 130 METHYLATION (TRI-) (BY SIMILARITY).
 FT MOD RES 552 552 METHYLATION (TRI-) (BY SIMILARITY).
 FT MOD RES 756 756 METHYLATION (TRI-) (BY SIMILARITY).
 SQ SEQUENCE 1938 AA; 223064 MW; D9A8A2EC5B182626 CRC64;

Query Match 9.2%; Score 249; DB 1; Length 1938;

Best Local Similarity 21.2%; Pred. No. 0.005;
 Matches 113; Conservative 80; Mismatches 179; Indels 160; Gaps 16;

QY 33 SQAPKEGAQARTAGALRDVSELSQLEDILSTYCVNNQGGPGDAGCEFAEPE 92
 Db 1091 SNLQSKIIEDEQALAMQ---LQKKIKELQARIIEEL-----EEIEAEASRA 1133
 QY 93 DAEKSTTYVARNGEPEPTPVVYGEKPSKGPDPNTEIROQ-DEVDGRDHRFPQKKKAG 151
 Db 1134 KAEKQSDLSR-----ELEISERLESEAGGATSAQIEMNKKREA 1172
 QY 152 LGRKBITLMTLNTLSTPEKLAALCK-----YALLEHRNSQOMKLLQKQSLVQ 206
 Db 1173 ---EFQMRDLREATLQHEATAATLRKHAQSVABLGQIDNLQVYKQKLEKSELKM 1229
 QY 207 EKDHLEGEHSKAVLARSKLESCLREL-----QRNRSLEKEGVQAR- 248
 Db 1230 EIDDLASNMETVSKAGNLEKMCRTLEDQVSELKTYEHEHORLINDLSAQRARLOTESGE 1289
 QY 249 -----EERKEKVTSH-FQVTLNDIQLQMBQ 274
 Db 1290 FSRQDEKDSLVQSLRGKQAFQIIEELKQLEESIKAKSALAHQASARHCDLLREQ 1349
 QY 275 HNER-----NSKLR-----QENVELAERLKKLIEOVELREEHIDK 309
 Db 1350 YEEQEAXABLRAMSKANSEVAQWTKYETDAIQRTTELEBAKQLAQRLQDAEHV-- 1407
 QY 310 VFKHQDLQOQLVDKLAQQAQOQMLKEABERHOREKDFLLKEAVESQRMCELMQOETHLKQ 369
 Db 1408 -----EAVNAKAS-----LEKTKRQLQNEVEDLMIDVERTNAACAALDKQRNFDK 1454
 QY 370 QLAITYKPEFPQNTLSKSSEVFTTFQEMEKTKKIKLEKETMTYRSWESSNKALLE 429
 Db 1455 ILAEWKHKYETH-----ALEASQKESRSLSSTEVFKVKNAYEESLDQLET 1500
 QY 430 MAEKTVRDRELGL--QV-----KIQRIEKLKCALQTERNDLNKRVQDLSA 474
 Db 1501 LKRENKLQGEISDLTQIAEGGKRIHELEKVKQVEQKSELQAALAEABA 1552

RESULT 11

REST_HUMAN
 ID REST_HUMAN STANDARD; PRT; 1427 AA.
 AC P30622;
 DT 01-APR-1993 (Rel. 25, Created)
 DT 01-APR-1993 (Rel. 25, Last sequence update)
 DT 28-FEB-2003 (Rel. 41, Last annotation update)
 DE Restin (Cytoplasmic linker protein-170 alpha-2) (CLIP-170) (Read-
 DE Sternberg intermediate filament associated protein).
 GN RSN.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 PP SEQUENCE FROM N.A.
 RC TISSUE=Peripheral blood monocytes;
 RX MEDLINE=92289675; PubMed=1600942;
 RA Bilbe G., Delabie J., Brueggem J., Richener H., Asselbergs F.A.M.,
 RA Carletti N., Sorg C., Odink K., Tarcay L., Wiesendanger W.,
 RA de Wolf-Petersen C., Shipman R.;
 RT "Restin: a novel intermediate filament-associated protein highly
 RT expressed in the Read-Sternberg cells of Hodgkin's disease.";

```

RL  EMBL J. 11:2103-2113 (1992).
RN  [2]
RP  SEQUENCE FROM N.A.
RX  MEDLINE=92405160; PubMed=1356075;
RA  Pierre P., Scheel J., Rickard J.E., Kreis T.E.;
RT  "CLIP-170 links endocytic vesicles to microtubules.";
RL  Cell 70:887-900(1992).
CC  -1- FUNCTION: SEEMS TO BE A INTERMEDIATE FILAMENT ASSOCIATED PROTEIN
CC  THAT LINKS ENDOCYTIC VESICLES TO MICROTUBULES.
CC  -1- SUBCELLULAR LOCATION: CYTOPLASMIC, ASSOCIATED WITH THE
CC  CYTOSKELETON.
CC  -1- ALTERNATIVE PRODUCTS:
CC  Event=Alternative splicing; Named isoforms=2;
CC  Name=Long;
CC  IsoId=P30622-1; Sequence=Displayed;
CC  Name=Short;
CC  IsoId=P30622-2; Sequence=VSP_000765;
CC  -1- TISSUE SPECIFICITY: HIGHLY EXPRESSED IN THE REED-STERNBERG CELLS
CC  OF HODGKIN'S DISEASE.
CC  -1- SIMILARITY: Contains 2 CAP-Gly domains.
CC  This SWISS-PROT entry is copyright. It is produced through a collaboration
CC  between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC  the European Bioinformatics Institute. There are no restrictions on its
CC  use by non-profit institutions as long as its content is in no way
CC  modified and this statement is not removed. Usage by and for commercial
CC  entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC  or send an email to license@isb-sib.ch).
CC  -----
DR  EMBL; X64838; CAA46050.1; -;
DR  EMBL; M97501; AAA35693.1; -;
DR  PIR; S22695; S22695.
DR  Genew; HGNC:10461; RSN.
DR  MIM; 179838; -.
DR  GO; GO:0005768; C:endosome; TAS.
DR  GO; GO:0005882; C:intermediate filament; TAS.
DR  GO; GO:0015630; C:microtubule cytoskeleton; TAS.
DR  GO; GO:0008017; F:microtubule binding; TAS.
DR  GO; GO:0006899; P:nonselective vesicle transport; TAS.
DR  InterPro; IPR00938; CAP-Gly.
DR  InterPro; IPR01878; Znf_CCHC.
DR  Pfam; PF01302; CAP GLY; 2.
DR  SMART; SM00343; Znf_C2HC; 1.
DR  PROSITE; PS00845; CAP_GLY_1; 2.
DR  PROSITE; PS0245; CAP_GLY_2; 2.
DR  Cytoskeleton; Microtubule; Coiled coil; Repeat; Alternative splicing.
KW  Cytoskeleton; Microtubule; CAP-GLY 1.
FT  DOMAIN 78 120
FT  SER-RICH.
FT  DOMAIN 143 204
FT  CAP-GLY 2.
FT  DOMAIN 232 274
FT  SER-RICH.
FT  DOMAIN 304 331
FT  COILED COIL (POTENTIAL).
FT  DOMAIN 350 1342
FT  CCHC-BOX.
FT  DOMAIN 1408 1421
FT  Missing (in isoform Short).
FT  VARSPLIC 457 491
FT  /FTid=VSP_000765.
FT  D -> B (IN REF. 2).
FT  CONFLICT 1069 1069
FT  SEQUENCE 1427 AA; 160989 MW; 0A4F166DD94254E8 CRC64;

Query Match 9.1%; Score 246; DB 1; Length 1427;
Best Local Similarity 24.4%; Pred. No. 0.0049;
Matches 114; Conservative 84; Mismatches 152; Indels 118; Gaps 20;

QY 117 KEPSKGDNTBEIRQSDVGR-----DHRPQKKKAGKLGK-----ITLLMOTLN 164
DQ 1770 KASSEGKEMKRLQOLEAABQIKHLBIENKNAESSKASSITRELOQRELKLTNLENLS 829
QY 165 TLS-----TPPEKLAICKYYAELLE-----HNSQOMKLLQKKQSOL----- 204
DQ 830 EYSQVKTLEKLEQLIKKFAEASVSVQSMQETVKNLHQKEEQNMLSSLEKLE 889
QY 205 -----VQEKHLRGEHSKAVLARSKLESICRELQR-----HNRS-----LKEEG 243
DQ 890 NLADMEAKFREKDS---REEQLIKAKLENDIAIMKMSGDSNLSQTKMNDLRLKRD 946

```

```

QY 244 VQAREEERKKEVTSHFQVTLNDTOLQWQ-HNERNKLRQENNELAERLKLIEQVEL 302
DQ 947 VEELQLKLTKANENASFLQKSIDMTVKAESQQAQKHEEKEKBLERKSLDEKQMET 1006
QY 303 REEHIDKV-----FKKD-----LQQLVDA--KLQAQAB-----MLKRAEE-RH 339
DQ 1007 SHNQOELKARYERATSETKYHEILQNLQKTLDTEDKLKAGARENSGLLQELERLK 1066
QY 340 QREKDFLKEAVESQRMCELMKQO-----ETHLKOQLALYTKPEEFQNT--L 385
DQ 1067 QADKAKAAQTAEDAMQIMEQTKETETILASLEDTKQTNKLQNLDTLTKENLNKVBEL 1126
QY 386 SKSSSEVFTTFKQEMKWTKKIKLEKETT-----MYRGRWESSNKA 426
DQ 1127 NKSGLLFTVENQKMBEERKEIETLKQAAQKQQLSALQENVKLAELGRSDEVTSHQ 1186
QY 427 LLEMAEETVRDKEGLQVKTQRLKLCRALQTERNDLNKRVQDLISA 474
DQ 1187 KLE--EERSVLNQL--LEMK-KRBSKFKDADEEKASLQKSITS 1229

RESULT 12
MYH8_HUMAN STANDARD; PRT; 1937 AA.
ID AC PI3535; Q14910;
DT 01-JAN-1990 (Rel. 13, Created)
DT 15-JUL-1998 (Rel. 36, Last sequence update)
DT 15-MAR-2004 (Rel. 43, Last annotation update)
DE Myosin heavy chain, skeletal muscle, perinatal (MyHC-perinatal).
GN MYH8.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX TISSUE=Skeletal muscle;
RC MEDLINE=90323631; PubMed=2373371;
RA Karsch-Mizrahi I., Peghali R., Shows T.B. Jr., Leinwand L.A.;
RT "Generation of a full-length human perinatal myosin heavy-chain-
RT encoding cDNA";
RL Gene 89:289-294(1990).
RN [2]
RP SEQUENCE FROM N.A.
RX TISSUE=Skeletal muscle;
RC MEDLINE=95324556; PubMed=7601129;
RA Jullian E.H., Kelly A.M., Pompidou A.J., Hoffman R., Schiaffino S.,
RA Stedman H.H., Rubinstein N.A.;
RT "Characterization of a human perinatal myosin heavy-chain
RT transcript.";
RL Eur. J. Biochem. 230:1001-1006(1995).
RN [3]
RP SEQUENCE OF 502-1937 FROM N.A.
RX TISSUE=Skeletal muscle;
RC MEDLINE=90235862; PubMed=1691980;
RA Rober E., Buchberger-Seidl A., Braun T., Singh S., Goedde H.W.,
RA Arnold H.H.;
RT "Identification of three developmentally controlled isoforms of human
RT myosin heavy chains.";
RL Eur. J. Biochem. 189:55-65(1990).
RN [4]
RP SEQUENCE OF 860-1937 FROM N.A.
RX MEDLINE=89234168; PubMed=2715179;
RA Peghali R., Leinwand L.A.;
RT "Molecular genetic characterization of a developmentally regulated
RT human perinatal myosin heavy chain.";
RL J. Cell Biol. 108:1791-1797(1989).
RN [5]
RP SEQUENCE OF 1-46 FROM N.A.
RA Esser K., Tidhar A., Myszowski M.;
RT "Isolation and characterization of the human perinatal MHC promoter.";
RL Submitted (MAY-1998) to the EMBL/GenBank/DBJ databases.
CC -!- FUNCTION: Muscle contraction.

```


-1- MISCELLANEOUS: Each myosin heavy chain can be split into 1 light meromyosin (LMM) and 1 heavy meromyosin (HMM). It can later be split further into 2 globular subfragments (S1) and 1 rod-shaped subfragment (S2).

-1- SIMILARITY: Contains 1 myosin-like globular head domain.

-1- SIMILARITY: Contains 1 IQ domain.

THIS SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

EMBL; V00430; CAA23712.1; --
 EMBL; J02714; AAA48972.1; --
 PIR; A29320; A29320.
 HSP; P13538; 2MYS.
 InterPro; IPR000048; IQ region.
 InterPro; IPR001609; myosin head.
 InterPro; IPR004009; myosin_N.
 InterPro; IPR002928; myosin_tail.
 Pfam; PF00612; IQ; 2.
 Pfam; PF00063; myosin_head; 1.
 Pfam; PF02736; Myosin_N; 1.
 Pfam; PF01576; Myosin_Tail; 1.
 PRINTS; PR00193; MYOSINHEAVY.
 ProDom; PD000355; myosin_head; 1.
 SMART; SM00015; IQ; 1.
 SMART; SM00242; MYSC; 1.
 PROSITE; PS00096; IQ; 1.
 Myosin; Muscle protein; Coiled coil; Thick filament; Actin-binding; Calmodulin-binding; ATP-binding; Methylation; Multigene family.

DOMAIN 1 785
 786 815
 844 1940
 COILED COIL (POTENTIAL).

NP_BIND 179 186
 ATP.

DOMAIN 660 682
 ACTIN-BINDING.

DOMAIN 762 776
 ACTIN-BINDING.

MOD_RES 130 130
 METHYLATION (TRI-) (POTENTIAL).

VARIANT 379 379
 G -> D.

CONFLICT 1547 1547
 T -> A (IN REF. 2).

CONFLICT 1913 1915
 ERA -> GRT (IN REF. 2).

SEQUENCE 1940 AA; 222816 MW; C34833D75B04DF2 CRC64;

Query Match 9.1%; Score 245; DB 1; Length 1940;
 Best Local Similarity 21.8%; Pred. No. 0.0072;
 Matches 114; Conservative 87; Mismatches 180; Indels 142; Gaps 16;

33 SCAPKPCGAQTAQSCALRDVSEELSRQLEDILSTVCVDNNGGPGEDGAGGEPAPPE 92
 1093 SQIOSKIEDEQALGNQ---LQKKIKELQARIEEL-----EEIEAERTSA 1135

93 DAEKSRVTYVANGPEPTPVVYGGKESKGPENTEEIRQSDVEGDRHRREQKKAGL 152
 1136 KAEKTRADLSRELE-----ISERLEAGGNTAAQI-----DMNKGREAFQKVRD 1182

153 GKEITLMTNTLSTPEKLAALCKY-----AELLEHRNSQKMLQKKQSQLYQEK 208
 1183 LEEATL-----QHEATAALRKGHADSTADVGEQIDLQVRQKLEKSEKSELQMEI 1233

209 DHLRGESKAVLARSKLSREL-----QRNRSLEEGVQVQAR-----248
 1234 DLLASNMESVSKANLEKYCRSLDQLSIKTKREBQORTINDISAKRLQTESGEYS 1293

249 -----EEBKKEVTSF-FQVTLNDIQLOMEOHN 276
 1294 RQVEKDALISQSRGKQAFQIQIEELKRHLEETKAKKCPAHALQSARHDCDLLREQYE 1353

277 ER-----NKKLR-----QENMELAEELKLEQVELREGEHIDKVP 311
 1354 EEQKAGELQRLSVANSEVAQWTKYETDALQRTTELEBAKKLQALQDAEEHV-----1409

QY 312 KHKDLQQLVDAKLQQAQSMLEKAEERHQEKDFLLKEAVESQRCMLAKQOETHLKOOL 371
 DB 1410 -----EAVNSKCA-----LEKTKQRLQNEVEDLMIDVERSNAAALDKQKDFKIL 1458

QY 372 ALYTEKFEFFONTLSKSEVFTTFQOEMKMTKKIKLSEKTTMYRSRWESSNKALLEWA 431
 DB 1459 SEWKQVEYETAELASQKESLSLSTELPKM-----KNAYEESLDHLLETUKRENKILQOEI 1514

QY 432 ESKTVRDKLEGLQVQIKQIKLQKLCALQTERNDLNKRVQDLQA 474
 DB 1515 SDLT---EQIAGGKA-IHELEKVKQIQEOKSELOTALSEAEA 1554

RESULT 14

RRBL HUMAN STANDARD: PRT; 1410 AA.
 ID RRBL HUMAN AC Q9P2E9; O75300; O968S2; Q9BWPI; Q9H476;
 DT 28-FEB-2003 (Rel. 41, Created)
 DT 28-FEB-2003 (Rel. 41, Last sequence update)
 DT 15-MAR-2004 (Rel. 43, Last annotation update)
 DE Ribosome-binding protein 1 (Ribosome receptor protein) (180 kDa
 DE Ribosome-binding protein) (ES/130 related protein).
 GN RRBPI OR KIAA1398.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 OK NCBI_TaxID=9606;
 RN [1]
 RP MEDLINE=98290552; PubMed=9628588;
 RA Langley R, Leung E, Morris C, Berg R., McDonald M., Weaver A.,
 RA Parry D., Ni J., Su J., Gentsz R., Spour N., Krissansen G.W.;
 RT "Identification of multiple forms of 180-kDa ribosome receptor in
 RT human cells.";
 RL DNA Cell Biol. 17:449-460(1998).
 RN [2]
 RP SEQUENCE FROM N.A. (ISOFORM 1 AND 2).
 RX MEDLINE=21638749; PubMed=11780052;
 RA Deloukas P., Matthews L.H., Ashurst J., Burton J., Gilbert J.G.R.,
 RA Jones M., Stavrides G., Almeida J.P., Babbage A.K., Bagguley C.L.,
 RA Bailey J., Barlow K.F., Bates K.N., Beard L.M., Beare D.M.,
 RA Beasley O.P., Bird C.F., Blakey S.E., Bridgeman A.M., Brown A.J.,
 RA Buck D., Burrill W.D., Butler A.P., Carder C., Carter N.P.,
 RA Chapman J.C., Clamp M., Clark G., Clark L.N., Clark S.V., Clee C.M.,
 RA Clegg S., Cobley V.E., Collier R.E., Connor R.E., Corby N.R.,
 RA Coulson A., Coville G.J., Deadman R., Dhali P.D., Dunn M.,
 RA Ellington A.G., Frankland J.A., Fraser A., French L., Garner P.,
 RA Grafham D.V., Griffiths C., Griffiths M.N.D., Gwilliam R., Hall R.E.,
 RA Hammond S., Harley J.L., Heath P.D., Ho S., Holden J.L., Howden P.J.,
 RA Huckle E., Hunt A.R., Hunt S.E., Jekosch K., Johnson C.M., Johnson D.,
 RA Kay M.P., Kimberley M.H., Leverhwa M.A., Lloyd C., Lloyd D.M., Lovell J.D.,
 RA Leveslaiho M.H., Levensha M.A., Lloyd C., Lloyd D.M., Lovell J.D.,
 RA Marsh V.L., Martin S.L., McConachie L.J., McKay K., McKerron A.A.,
 RA Milne S.A., Mistry D., Moore M.J.F., Mullikin J.C., Nickerson T.,
 RA Oliver K., Parker A., Patel R., Pearce T.A.V., Peck A.I.,
 RA Phillimore B.J.C.T., Prathalingam S.R., Plumb R.W., Ramsay H.,
 RA Rice C.M., Ross M.T., Scott C.E., Sehra H.K., Showkeen R., Sims S.,
 RA Skuce C.D., Smith M.L., Soderlund C., Steward C.A., Sulston J.E.,
 RA Swann R.M., Sycamore N., Taylor R., Tee L., Thomas D.W., Thorpe A.,
 RA Tracey A., Tromans A.C., Vaudin M., Wall M., Wallis J.M., Thorpe A.,
 RA Whitehead S.L., Whittaker P., Willey D.L., Williams L., Williams S.A.,
 RA Wilming L., Wray P.W., Hubbard T., Durbin R.M., Bentley D.R., Beck S.,
 RA Rogers J.;

	POTENTIAL.
CYTOPLASTIC (POTENTIAL).	
21 X 10 AA APPROXIMATE TANDEM REPEATS OF	
N-Q-G-K-A-E-G-A-P.	
Missing (in isoform 2).	
/FTIG-VSP 003949.	
Missing (in isoform 1 and isoform 2).	
/FTIG-VSP 003950.	
G -> A (IN REF. 1; AAC25978).	
A -> G (IN REF. 1; AAC25978).	
K -> KTTAEAQGKKAEGAQIQKKNGAQTQGCKAKGA	
QNQKNBGAQTGGKABGAQTGGKADQAQQGKKBAGAQ	
NOGKBAEQGNQKKAEGANOGKADGANONQKKAEGAO	
OQSKABAQNOST (IN REF. 1; AAC25978).	
S -> A (IN REF. 1; AAC25978 AND 2).	
V -> A (IN REF. 1; AAC25978).	
Q -> QDKKEAGNQRKKAEGANOGRAGAENQOKKAAE	
GAPNQKKGAPNQKKALBTPTNQKKAETPNQKKKAE	
TYNQKKAEGANOQKKAEGANOQKTASGTTPNQKKAEGA	
NOQKKAEGANOQKKAEGANOQKKAEGANOQKKAEGAO	
(IN REF. 2).	
L -> H (IN REF. 1; AAC25977, 2 AND 4).	
S -> L (IN REF. 2 AND 4).	
T -> M (IN REF. 1; AAC25977).	
Query Match Score 244.5; DB 1; Length 1410;	
Best Local Similarity 22.4%; Pred. No. 0.0056;	
Matches 151; Conservative 109; Mismatches 237; Indels 177; Caps 32;	
QY 1 KSPPGPAG--PEGNAERPSOAAAFAVB--AESPGSSQAPRPEGEQAATAQ-----48	
DB 480 KKAEAQNQKKKAEGANOQKKAEGANOQKGGBEAONOQKTEGTAQGGKAERSPNQGXK 539	
QY 49 -----SGAIRDVSELSRLQLDILLSTYCVDNNQGGPEDGAQGESPAEPDAE----K 96	
DB 540 GEGAPIQGHKASVANQGTKVEGI-----TNQG---KKAEGSSSEGKKAESGPSNOGK 598	
QY 97 SRTTVARNQEPEPTPVVYGEEKPKSGKDPTBEIRQSDEVGDRHRRPQEK---KAG--151	
DB 589 KADAANAQNKTESAVQGR-----NT-DVAQSPFAPKHQ-APAXKKSGSXKKHGP 637	
QY 152 -----LGKETITLMQTLNLTSIPE---EKLAAUCKKYVALLEE-HRNKSQ-CMKL 196	
DB 638 GPDDADGPLYLPTYKTLVTGVSNVFNEGEARLLIELSEKAGIIQDTWHKATKGDPVAI 697	
QY 197 LAQKOSQLOVKEDHLHGHSKAVLARSLKLESCREL--ORHNRSLKIEGVQR---AREEE 251	
DB 698 LR--QLBEKEKLLATEQEDAIVAKSKLRILINKEMAEKAKAAAGEAKTKVLVAROE 754	
QY 252 -----EKEKVYSHFQVLTNDILOMQEO-HNRNSKLPQNMELAEPLIKLI 297	
DB 755 ITAVQARMQASTVRHVKEY-QOLQKRIYLOLENQPNTOLARLOQENSILEDALNQAT 813	
QY 298 EGYE-----LREEHIDRYFK-----HKDLQOOLLVDAKLQOAQMELKRAEB 338	
DB 814 SQVESKQNAELAKLROE-LSKVSKEIAVEKSEA VRODEQQRKALEAKAAAAFEKQVLAQAS 872	
QY 339 HQRKDFLIKHA VESORMCELMHQQBETHUKQOLALYTKEFPONTLS-----KSEVP 392	
DB 873 HRSEBALQRLDEVSR--ELCHTOSSH--ASLRADAERKAQGOQQOMAEILSHSKLOSSEAE 928	
QY 393 TTFK-----QEMEKMTKIKCLEKETTMVRSWESSNKALLEMAENTVR 437	
DB 929 VRSKCBELSGHLQOLEARAENSELTERIRS IEALLGAQARDADQVASQAEADQQOTR 988	
QY 438 DKELE---GLQVKIORBLCRALQATERNDLN----KRVDLSAGQG-----SLTDS 483	
DB 989 LKELESQVGSLEKATEIREAVEQQVKQNDLREKWNKAMEALATAEQACKELLILTQA 1048	
QY 484 GPERRP-----EGPGAQ-----PSSPRVTEAPCYCAPST 514	
DB 1049 KESEKQCLCIIRAQTYMELLAALLPELVUAQNYTWLOLDKEXGKITLKHPAPAEPSS 1108	


```
QY 515 EASGQTGPQBPSTA 528
Db 1109 DLASKLREABETQS 1122

RESULT 15
SCPI1 MOUSE
ID SCPI1_MOUSE STANDARD; PRT; 993 AA.
AC Q62209; 009205; P70192; Q62329;
DT 15-JUN-1998 (Rel. 36, last sequence update)
DT 15-JUL-1998 (Rel. 36, last sequence update)
DT 15-DEC-1998 (Rel. 37, last annotation update)
DE Synaptonemal complex protein 1 (SCP-1 protein).
GN SCPI1 OR SCPI1.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN (1)
RN SEQUENCE FROM N.A.
RC STRAIN=CBA; TISSUE=Testis;
RX MEDLINE=96004899; PubMed=7548215;
RA Sage J., Martin L., Cuzin F., Rassoulzadegan M.;
RF "cDNA sequence of the murine synaptonemal complex protein 1 (SCP1).";
RL Biochim. Biophys. Acta 1263:256-260(1995).
RN (2)
RN SEQUENCE FROM N.A.
RC STRAIN=Swiss; TISSUE=Testis;
RA Kerr S.M., Taggart M.H., Lee M., Cooke H.J.;
RL Submitted (APR-1995) to the EMBL/GenBank/DBAJ databases.
RN (3)
RN SEQUENCE OF 1-149 FROM N.A.
RC STRAIN=C57BL/6;
RA Sage J., Li Y., Martin L., Mattei M.-G., Guenet J.-L., Liu J.G.,
RL Hoog C., Cuzin F., Rassoulzadegan M.;
RL Submitted (JAN-1997) to the EMBL/GenBank/DBAJ databases.
RN (4)
RN SEQUENCE OF 95-787 FROM N.A.
RC STRAIN=ICR; TISSUE=Testis;
RA Tsuchida J., Nishina Y., Nozaki M., Uchida K., Nishimune Y.;
RL Submitted (NOV-1996) to the EMBL/GenBank/DBAJ databases.
CC -!- FUNCTION: Major component of the transverse filaments of
CC synaptonemal complexes (SCS), formed between homologous
CC chromosomes during meiotic prophase.
CC -!- SUBCELLULAR LOCATION: Nuclear. In tripartite segments of
CC synaptonemal complexes, between lateral elements in the nucleus.
CC Found only where the chromosome cores are synapsed. Its N-terminus
CC is found towards the centre of the synaptonemal complex while the
CC C-terminus extends well into the lateral domain of the
CC synaptonemal complex (By similarity).
CC -!- DOMAIN: Consists of an alpha-helical stretch of 700 AA residues,
CC flanked by N- and C-terminal globular domains. The C-terminal
CC domain has DNA-binding capacity (By similarity).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; Z381118; CAAS6262.1; -
CC EMBL; L41069; AAA64514.1; ALT_INIT.
CC EMBL; U62864; AAC53335.1; -
CC EMBL; U62860; AAC53335.1; JOINED.
CC EMBL; U62861; AAC53335.1; JOINED.
CC EMBL; U62862; AAC53335.1; JOINED.
CC EMBL; U62863; AAC53335.1; JOINED.
CC EMBL; D88539; BAAL3639.1; -
CC PIR; S49461; S49461.
CC MGD; MGI:105931; Sycp1.

GO; GO:0000795; C:synaptonemal complex; IDA.
DR InterPro; IPR008827; SCP-1.
DR Pfam; PF05483; SCP-1; 1.
KW Nuclear protein; Meiosis; Cell division; Phosphorylation;
KW DNA-binding; Coiled coil.
FT DOMAIN 12 97 ASP/GLU-RICH (ACIDIC).
FT DOMAIN 104 815 COILED COIL (POTENTIAL).
FT DOMAIN 114 117 NUCLEAR LOCALIZATION SIGNAL (POTENTIAL).
FT DOMAIN 697 700 NUCLEAR LOCALIZATION SIGNAL (POTENTIAL).
FT DOMAIN 898 901 NUCLEAR LOCALIZATION SIGNAL (POTENTIAL).
FT DOMAIN 978 986 ARG/LYS-RICH (BASIC).
FT CONFLICT 527 527 F -> L (IN REF. 2).
SQ SEQUENCE 993 AA; 115962 MW; 1A4FA790D64FAF66 CRC64;

Query Match 9.0%; Score 243.5; DB 1; Length 993;
Best Local Similarity 22.9%; Pred. No. 0.0043;
Matches 124; Conservative 91; Mismatches 183; Indels 143; Gaps 23;

QY 34 QAPRKEGGAQAPTAQSGALRDVSELSROLED-----ILSYCVND 73
Db 132 QENRKLIIEAQRKAIQI--LQFENEKYSKLKEEIQENKDLKENNATIEWCNLLKFTC-- 187
QY 74 NNQGFGEDEGAQGEPAEPEDAERKRTYVARNGEPPTVPVYGEKPSKGDPTNTEERQSD 133
Db 188 -----ARSAEKTNKYEERETRQVYVDINSNIEMILAF-----EELRVQA 229
QY 134 EVG-----DRDHERPQ--EKKKAGLGKKEITLLMOTLTLSTPBEKLAALCKKYAEL 183
Db 230 ENARLEHGFKLKEDHEKIQHLEBEYQKRVNNKQVSELLIQSAEKENM-----KDLTFL 285
QY 184 LEEHRNSQKQW----KLLQKKQSQLVQEKDHLRGEHSKAVLARSKLESLECRELQRH---N 236
Db 286 LEESRDKANQLEKTKLQDENLKLSEKNDHL-----TSELEDIKMSQRMSTQ 335
QY 237 RSLKEE-----GVQARABEEERKEVTSHPV-----TLNDIQLOMEQHN 276
Db 336 KALEEDLIQATKTSIQLTVEYKRAQMEELNKAKTHTSFVVTTELKATCTTSELRLTEQRL 395
QY 277 ERNS-----KLROENMELAEELKXLIQYELREBEHIDKVFKHQLOQLOQVADKLQQA 328
Db 396 EKEDQLKLITVELQKSNEL--EEMTKFKNKVELEELKNILAE---DQKLIDEK-KQV 450
QY 329 QEMLKAEERHQRKDFLL-----KRAVESQRMCEKAKQOETH-LKQOLALYTE-KKEEFQ 382
Db 451 EKLAELQEKQ--ELTFLLETREREVDLQEQVTVTKTSBQHYLKQVEEMKTELEKELK 509
QY 383 NT-----LSKSSEVFTTFKQEMKMTTKTKLEKETTMY 416
Db 510 NTELTASCDMLLENKFFVQESADMALELKKHQEDIIINCKQERLLKQIENLEEKEMHL 569
QY 417 RSRWESNKALLEWAER-KTVRDKELF---GLQVKQRLEKLCALQTERNDLNKRVQDL 472
Db 570 RDELESVRKEFIOQGDVEVKCKLDKSEENARSIECEVLKKEQMKILESKNNLKQVENEK 629
QY 473 S 473
Db 630 S 630

RESULT 16
PLE1 CRIGR
ID PLE1 CRIGR STANDARD; PRT; 4473 AA.
AC Q9J155;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, last sequence update)
DT 15-MAR-2004 (Rel. 43, last annotation update)
DE Plectin 1 (PLTN) (PCN) (300-kDa intermediate filament-associated
DE protein) (IPAP300) (Fragment).
GN PLE1
OS Cricetus griseus (Chinese hamster),
OC Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Cricetinae;
OC Cricetulus.
```


RA SODA M., Misumi Y., Fujiwara T., Nishioka M., Ikehara Y.;
 RT "Molecular cloning and sequence analysis of a human 372-kDa protein
 RL localized in the Golgi complex";
 CC Blochem. Biophys. Res. Commun. 205:1399-1408(1994).
 CC -!- FUNCTION: May participate in forming intercisternal cross-bridges
 CC of the Golgi complex.
 CC -!- SUBUNIT: Homodimer; disulfide-linked.
 CC -!- SUBCELLULAR LOCATION: Golgi; membrane-associated.
 CC -!- DISEASE: Antigen in chronic rheumatoid arthritis and in the
 CC autoimmune disease Sjogren's syndrome.
 CC -!- SIMILARITY: Belongs to the golgin family.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC

EMBL: X75304; CAA53052.1; --
 EMBL: D25542; BAA05025.1; --
 PIR: A56539; A56539.
 PIR: I52300; I52300.
 DR Genew; HGNC:4429; GOLGB1.
 DR MW: 602500; --
 DR GO: GO:0000139; C:Golgi membrane; TAS.
 DR GO: GO:0005795; C:Golgi stack; TAS.
 DR GO: GO:0016021; C:integral to membrane; TAS.
 DR GO: GO:0007030; P:Golgi organization and biogenesis; TAS.
 KW Golgi stack; Antigen; Coiled coil; Transmembrane.
 FT DOMAIN 1 3235 CYTOPLASMIC (POTENTIAL).
 FT TRANSMEM 3236 3256 POTENTIAL.
 FT DOMAIN 3257 3259 LUMENAL (POTENTIAL).
 FT DOMAIN 48 593 COILED COIL (POTENTIAL).
 FT DOMAIN 677 1028 COILED COIL (POTENTIAL).
 FT DOMAIN 1062 1245 COILED COIL (POTENTIAL).
 FT DOMAIN 1301 1779 COILED COIL (POTENTIAL).
 FT DOMAIN 1828 3185 COILED COIL (POTENTIAL).
 FT DOMAIN 2420 2423 POLY-GLU.
 FT DOMAIN 2993 2996 POLY-SER.
 FT CONFLICT 1 39 MISSING (IN REF. 3).
 FT CONFLICT 215 215 A -> AQLSSM (IN REF. 3).
 FT CONFLICT 1765 1765 D -> G (IN REF. 3).
 FT CONFLICT 2950 2950 H -> D (IN REF. 3).
 SQ SEQUENCE 3259 AA; 376075 MW; 60376A20D8A178DD CRC64;

Query Match 9.0%; Score 242; DB i; Length 3259;

Best Local Similarity 20.4%; Pred. No. 0.016;

Matches 121; Conservative 99; Mismatches 168; Indels 206; Gaps 22;

QY 43 QARTAGSALRDVSELSRQLEDILSTYC-----VDNNGGPGEDGAGGPAPBDEAK 96
 DB 1743 QSLMSEKQSLSEEVQDLKHQIEDNVSKQANLEATEKHNDQNTVTEEGTQSPGTEGDS 1802
 QY 97 -STYVARGPEPT-----PVYGEKPSKGDPTNTEIROSDVGDHRHRRPOEKKAK 150
 DB 1803 LMSSTRPTCSVPYSAKSNAPV--SKDPSSHDEINNYLIQDQLKRIAGLEBEKQKNK 1860
 QY 151 GLCKEI-----TLTMO-----TLNTLS-----TPBE--- 171
 DB 1861 EFSQTLNENKNTLLSQISIKGELKMLQBEVTKMLNLLNQIQEELSRVTKLKTAREBKD 1920
 QY 172 -----KLAL-----CKYYA-----ELLE-EHRNSKQMKLLQKQSOQLVBEKD 209
 DB 1921 DLBERLNLQALNLSIGNYCDQVDTAQIKNLESEMKNLKKCVSELEBEKQQLVKEKT 1980
 QY 210 HLRGHSKAVLAR-----SKLSLCELRQHRNSLK----- 240
 DB 1981 KYSEIRKYLEKIQAQKEPGNKSNAKLEQLLKEKQEVKQLQKDCIRYQEKISALER 2040
 QY 241 -----EGVQARAREERKRVKVTSHQVTLNDIQLQEQHQRNENSK 281

DB 2041 TVKALEFVOTESQOLEITKENLAQAVEHKKQAELASFKVLLDDTQSEARVLADNLK 2100
 QY 282 LRQENNELASRLKGLILOYELREHHDVKFKKOLQOOLVDAKLQQAQEMLKEASEHQR 341
 DB 2101 LKGLQSNKESVKSQWKQ-----KQEDLERR-----LQAAEKH-- 2134
 QY 342 EKDFLLKAVESQRMCELMKQQRTHUKQALALYTEKFEFQNTLSKSSSEVFTTFKQEMEK 401
 DB 2135 -----LKEKQNMQEKLDALREKVLHEETIG-----BIQVTLNKKCKQEVQQLQENLDS 2182
 QY 402 -----MTKKIKKLEKETTY-----RSRWESSNKALLEMAEERTVDKE----- 440
 DB 2183 TVTQLAAPTYSMSLQDDRDRVIDEAKKWERKPSDAIQSEEB-ILKEDNCSVLKQDLR 2241
 QY 441 -----LEGQVQKIQRL-----KLCRALQTERNDLAKRVQD 471
 DB 2242 QMSIRMEELKINISRLHDKQIWESRAQTEVQLOQKVCDTLQENKELLSQLEE 2295

RESULT 19

CENE_HUMAN

ID CENE_HUMAN STANDARD; PRT; 2663 AA.

AC Q02224;
 DT 01-JUL-1993 (Rel. 26, Created)
 DT 01-JUL-1993 (Rel. 26, Last sequence update)
 DT 10-OCT-2003 (Rel. 42, Last annotation update)
 DE Centromeric protein 2 (CENP-E protein).
 GN CENPE
 OS Homo sapiens (Human)
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA MEDLINE=93024922; PubMed=1406971;
 RA Yen T.J., Li G., Schaar B.T., Szilak I., Cleveland D.W.;
 RT "CENP-E is a putative kinetochore motor that accumulates just before
 RT mitosis.";
 RL Nature 359:536-539(1992).
 RN [2]
 RP CHARACTERIZATION.
 RA MEDLINE=95196755; PubMed=7889940;
 RA Thresher D.A., Jordan M.A., Schaar B.T., Yen T.J., Wilson L.;
 RT "Mitotic HeLa cells contain a CENP-E-associated minus end-directed
 RT microtubule motor.";
 RL EMBO J. 14:918-926(1995).
 RN [3]
 RP CHARACTERIZATION.
 RA MEDLINE=98437347; PubMed=9763420;
 RA Chan G.K.T., Schaar B.T., Yen T.J.;
 RT "Characterization of the kinetochore binding domain of CENP-E reveals
 RT interactions with the kinetochore proteins CENP-F and HUBB1.";
 RL J. Cell Biol. 143:49-63(1998).
 RN [4]
 RP FARNESYLATION.
 RA MEDLINE=20459117; PubMed=10852915;
 RA Ashar H.R., James L., Gray K., Carr D., Black S., Armstrong L.,
 RA Bishop W.R., Kirschmeier P.;
 RT "Farnesyl transferase inhibitors block the farnesylation of CENP-E
 RT and CENP-F and alter the association of CENP-E with the
 RT microtubules.";
 RL J. Biol. Chem. 275:30451-30457(2000).
 CC -!- FUNCTION: MINUS-END DIRECTED MICROTUBULE MOTOR. PROBABLE
 CC KINETOCHORE MOTOR. ACCUMULATES JUST BEFORE MITOSIS AT THE G2 PHASE
 CC OF THE CELL CYCLE. PROBABLY IMPORTANT FOR CHROMOSOME MOVEMENT
 CC AND/OR SPINDLE ELONGATION.
 CC -!- SUBUNIT: INTERACTS WITH CENP-F AND HUBB1 KINASE.
 CC -!- SUBCELLULAR LOCATION: ASSOCIATES WITH KINETOCHORES DURING
 CC CONGRESSION, RELOCATES TO THE SPINDLE MIDZONE AT ANAPHASE, AND IS
 CC QUANTITATIVELY DISCARDED AT THE END OF THE CELL DIVISION.
 CC -!- SIMILARITY: Belongs to the kinesin-like protein family.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).

DR EMBL: Z15005; CAA78727.1; --
DR PIR: S28261; S28261.
DR HSSP: P17119; 3XAR.
DR Genew: HGNC:1856; CENPE.
DR GK: Q02224; --
DR MIM: 117143; --
DR GO: GO:0005699; C:kinetochore; TAS.
DR GO: GO:0005334; C:nucleus; TAS.
DR GO: GO:0008350; F:kinetochore motor activity; TAS.
DR GO: GO:0000067; P:DNA replication and chromosome cycle; TAS.
DR GO: GO:0000709; P:mitotic chromosome movement; TAS.
DR GO: GO:0000780; P:mitotic metaphase plate congression; TAS.
DR InterPro: IPR001752; kinesin_motor.
DR Pfam: PF00225; kinesin; 1.
DR PRINTS: PR00380; KINESINHEAVY.
DR SMART: SMO0129; KISC; 1.
DR PROSITE: PS00411; KINESIN MOTOR DOMAIN1; 1.
DR PROSITE: PS00067; KINESIN MOTOR DOMAIN2; 1.
DR Motor protein; Cell division; ATP-binding; Coiled coil; Mitosis;
KW Cell cycle; Centromere; Lipoprotein; Prenylation.
FT DOMAIN 1 335 KINESIN-MOTOR.
FT DOMAIN 336 2471 COILED COIL (POTENTIAL).
FT DOMAIN 2472 2663 GLOBULAR (POTENTIAL).
FT NP_BIND 86 93 ATP (BY SIMILARITY).
FT LIPID 2660 2660 S-farnesyl cysteine.
SQ SEQUENCE 2663 AA; 312087 MW; CFCF1380C8C8CB8 CRC64;

Query Match 8.9%; Score 240; DB 1; Length 2663;

Best Local Similarity 20.2%; Pred. No. 0.015;

Matches 130; Conservative 101; Mismatches 176; Indels 238; Gaps 25;

QY 55 VSELSRQLEDILSTYCVNNQGGPBGAGG-----EPASPDAAKRTVA----- 102
DB 987 ISESVRNLR-----MEETGETDEFOQKVGIDKKODLEAKNTQTLTADYKDNELI 1039
QY 103 -----RNGSEPTPVVYGEKPSGSD-----PNTETIR-QSDRVGDRDH 140
DB 1040 EQQRKIFSLIOEKNELOQMLESVTAEREQLTKDKNEMTIENTQEERLLGLDKKQOE 1099
QY 141 RPOEKKA-KGLGKEITLLMTLNTLSTPEKL----- 173
DB 1100 IVAQEKNAIKKEG-----LSRTCDRLAEVBEKLEKSKSQQLQEQQLNVOERMSEMQK 1155
QY 174 -----AALCKYAELEEHNSOKMKLQKQSQQLVQOE 207
DB 1156 KINEIENLKNELKNELTLEHMETERLELAQKLANEYEVKSIKRRKVLKELQKSPETS 1215
QY 208 XDLRLG-----EHSKAV--LARSKLESLC-----RELQRHN 236
DB 1216 RDHLGYREIETATGLTKELKLAHILKHEQTHDELRESVSEKTAQIINTQDLKSH 1275
QY 237 RSLKEGVQVQARBEKKKQV--TSHFQVTLNDIQLQWQHNRS-----KLRENMLA 290
DB 1276 TKLQBE-IPVLHEEQELLPNVKVSVETQETVNELELLTEQSTTDDSTTLARIEMERLRN 1334
QY 291 EELKXLIQYELREHIDKV-----FKHKLQQLQVLD--AKLQQAQ----- 329
DB 1335 EKFSQSEIKSLTKERONLTKIEALRVKDKQLKEHIRETLAKIQSSQKQESLNWKE 1394
QY 330 -----EML-----KEAEERHQREKDFLKEAVEQSRCMCEL 359
DB 1395 KDNETTKIVSEMQPKDSALLRIEIEMLGLSKRLQSHDEMKS-VAKEKXDLQLRQEV 1453
QY 360 MKQETHLKOOL-----ALYTEKFRFQ-----NTLSKSEVFTTFKQ 397
DB 1454 LOSESDQLKENIKETIAKHLETEBELKVAHCLKEQBETINELRVNLSEKETEISTQKQ 1513

QY 398 ---EMEKTKIKKL-EKETTM---YRSRWSSNKALLEMAEKTIVROKELGLOVKI-- 448
DB 1514 LEAINDKQNKQIQIYEKEEQNLNIQISEVQNVNNEKQFKEHRAKDSALQSISSZMLE 1573
QY 449 -----ORLEKLCRALQTERNDLNKRVQDLSA 474
DB 1574 LTNRLQESQEBTQIMIKKEENKRVQEAQIERDQKENTKEIVA 1618

RESULT 20

TRHY_HUMAN STANDARD; PRT; 1898 AA.
ID TRHY_HUMAN
AC Q07283;
DT 01-OCT-1994 (Rel. 30, Created)
DT 01-OCT-1994 (Rel. 30, Last sequence update)
DE 15-MAR-2004 (Rel. 43, Last annotation update)
DE Trichohyalin.
GN THH OR TRHY OR THH.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=93380194; PubMed=7685034;
RA Lee S.-C., Kim I.-G., Marekov L.N., O'Keefe B.J., Parry D.A.D.,
RA Steinert P.M.;
RT "The structure of human trichohyalin. Potential multiple roles as a
RT functional EF-hand-like calcium-binding protein, a cornified cell
RT envelope precursor, and an intermediate filament-associated (cross-
RT linking) protein";
RT J. Biol. Chem. 268:12164-12176(1993).
RN [2]
RP SEQUENCE OF 1731-1898 FROM N.A., AND CHARACTERIZATION.
RX MEDLINE=93315897; PubMed=7686953;
RA O'Keefe B.J., Hamilton E.H., Lee S.-C., Steinert P.M.;
RA "Trichohyalin: a structural protein of hair, tongue, nail, and
RT epidermis";
RL J. Invest. Dermatol. 101:655-715(1993).
CC -I- FUNCTION: Intermediate filament-associated protein that associates
CC in regular arrays with keratin intermediate filaments (KIF) of the
CC inner root sheath cells of the hair follicle and the granular
CC layer of the epidermis. It later becomes cross-linked to KIF by
CC isodipeptide bonds. It may serve as scaffold protein, together
CC with involucrin, in the organization of the cell envelope or even
CC anchor the cell envelope to the KIF network. It may be involved in
CC its own calcium-dependent postsynthetic processing during terminal
CC differentiation.
CC -I- SUBUNIT: Monomer (Probable).
CC -I- TISSUE SPECIFICITY: Found in the hard keratinizing tissues such as
CC the inner root sheath (IRS) of hair follicles and medulla, and in
CC the filiform papillae of dorsal tongue epithelium (Probable).
CC -I- DEVELOPMENTAL STAGE: Expressed during late differentiation of
CC the epidermis.
CC -I- DOMAIN: Consists of nine domains. Domain 1 contains two EF-hand
CC calcium-binding domains. Domains 2-4, 6, and 8 are almost
CC entirely alpha-helical, configured as a series of peptide repeats
CC of varying regularity, and are thought to form a single-stranded
CC alpha-helical rod stabilized by ionic interactions. Domain 6 is
CC the most regular and may bind KIF directly by ionic interactions.
CC Domains 5 and 7 are less well organized and may induce folds in
CC the molecule. Domain 9 contains the C-terminus, conserved among
CC different species.
CC -I- PFM: Substrate of transglutaminase. Some 200 arginines are
CC probably converted to citrullines by peptidylarginine deiminase.
CC -I- SIMILARITY: In the N-terminal section; belongs to the S-100
CC family.
CC -I- SIMILARITY: Contains 2 EF-hand calcium-binding domains.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its

use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

```
CC ENBL; L09190; AAA65582.1; -.
CC PIR; A45973; A45973.
CC HSP; P02633; 4ICB.
CC Genew; HGNC:11791; THH.
CC MIN; 190370.
CC GO; GO:0005856; C:cytoskeleton; NAS.
CC GO; GO:0005509; F:calcium ion binding; TAS.
CC InterPro; IPR001751; CaBP_S100.
CC InterPro; IPR002048; EF-hand.
CC InterPro; IPR002017; Spectrin.
CC Pfam; PF00036; efhand; 1.
CC Pfam; PF01023; S_100; 1.
CC ProDom; PD003407; CaBP_S100; 1.
CC PROSITE; PS00018; EF_HAND; 1.
CC PROSITE; PS00303; S100_CaBP; 1.
CC Keratinization; Calcium-binding; Repeat; Citrullination.
DOMAIN 1 91
CA_BIND 22 33
CA_BIND 62 73
CA_BIND 314 390
REPEAT 314 326
REPEAT 327 339
REPEAT 340 351
REPEAT 352 364
REPEAT 365 377
REPEAT 378 390
DOMAIN 391 444
REPEAT 391 444
REPEAT 397 402
REPEAT 403 408
REPEAT 409 414
REPEAT 415 420
REPEAT 421 426
REPEAT 427 432
REPEAT 433 438
REPEAT 439 444
DOMAIN 444 702
DOMAIN 923 1162
REPEAT 923 952
REPEAT 953 982
REPEAT 983 1012
REPEAT 1013 1042
REPEAT 1043 1072
REPEAT 1073 1102
REPEAT 1103 1132
REPEAT 1133 1162
DOMAIN 1250 1849
CONFLICT 1752 1782
CONFLICT 1794 1801
CONFLICT 1857 1857
CONFLICT 1880 1880
SEQUENCE 1898 AA; 247219 MM; A74B5947FB62E31D CRC64;
Query Match 8.9%; Score 239.5; DB 1; Length 1898;
Best Local Similarity 22.3%; Pred. No. 0.012;
Matches 125; Conservative 102; Mismatches 231; Indels 103; Gaps 21;
QY 8 EAGPEGAQRPQAPAVAEFGSSQAPRKPEGAQTAQSGALRDVSELSRQLEIIL 67
Db 262 EEPQQRLEQEEELKLEQLRRERQEEQEQQLRRRQQLRRRQEEERQEE-- 319
QY 68 STYVDNNGGPGEDGAQGEPAPEDAKSRITYVANGPEPTPVYGEKPSKDPNTE 127
Db 320 -----ERREQERQEE-ERREQERQEEERREQERQEEERQEEERRE---Q 362
QY 128 ETRQSDVEGDGDD---HRRFOEKKKAGLKGKTEILLMQTLNTLSTPEEKLAALCKKYABLL 184
```

```
Db 363 QLRRQEERREQLRRRQEEERREQLRRRQQL-----RREQQL-----RREQQLR 409
QY 185 EEH--RNSOKMKLQKQSLVQEKDHLRGHSHKAVLARSKLESCLRELQHNHSLKEE 242
Db 410 REQLRRRQQLRRRQEEERREQLRRRQQL-----LRREQEEE--RHEQKHEQERREQ 464
QY 243 GVOARAEHEE---KKEVTISHPOVTLNDIOLQMEQHNSKSL-----RQENMELAEKAK 295
Db 465 RLKREQEERDRLKKEHEETERHEQERQKQQLKRDQEEERERERWLKLEERREQLRRREQ 524
QY 296 LJ--EQLYELREHIDKVPFKDLOQL-----VDAKLQQAQMLKBAERHQ 340
Db 525 QLRRQEERREQLRRRQEEERREQLRRRQQLRRRQEEERLEQLLKREBEKLEQERREQ 584
QY 341 R-----BKDFLLAEVRSQRMCHMKQETHLQQLALYTEKPEEFQNTLSKSEVFT 393
Db 585 RLKREQEERDRLKKE--ERREQRLKREQEERLEQLK--REVERLEQE--RRDERLK 640
QY 394 TPKQEMKMTKKIKLEKETTMYRSRWSSNKALLEMAEKTVRDKLEGLQVKIOLK 453
Db 641 REPEERERHLLKSEQE-----ERHQLRRRQEEERREQLKREBEERLEQLK-- 693
QY 454 LCRALQTERNDLNKRVQDISAGQGSGLTSGPERPE-----GPGAQAPSSPRVTEAPCY 508
Db 694 -----EHEERREQLAEQEEOARERIKSRIPKQWQLESEADARQSKVLLR 743
QY 509 PGAPSTEASQGTGPQEPPTSAR 529
Db 744 -----QAGRAEAPQEEK 758
RESULT 21
MYH1_HUMAN STANDARD; PRT; 1939 AA.
AC P12882; Q9Y622;
DT 01-OCT-1989 (Rel. 12, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 15-MAR-2004 (Rel. 43, Last annotation update)
DE Myosin heavy chain, skeletal muscle, adult 1 (Myosin heavy chain
DE IIX/d) (MyHC-IIX/d).
GN MYH1.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OC NCBI_TaxID=9606;
[1]
SEQUENCE FROM N.A.
TISSUE=Skeletal muscle;
MEDLINE=99318869; PubMed=10388558;
RA Weiss A., Schiaffino S., Leinwand L.A.;
RT "Comparative sequence analysis of the complete human sarcomeric myosin
RT heavy chain family: implications for functional diversity.";
RL J. Mol. Biol. 290:61-75(1999).
[2]
SEQUENCE OF 1064-1939 FROM N.A.
RA MEDLINE=86176778; PubMed=2421254;
RA Saez L., Leinwand L.A.;
RT "Characterization of diverse forms of myosin heavy chain expressed in
RT adult human skeletal muscle.";
RL Nucleic Acids Res. 14:2951-2969(1986).
CC -!- FUNCTION: Muscle contraction.
CC -!- SUBUNIT: Muscle myosin is a hexameric protein that consists of 2
CC heavy chain subunits (MHC), 2 alkali light chain subunits (MLC)
CC and 2 regulatory light chain subunits (MLC-2).
CC -!- SUBCELLULAR LOCATION: Thick filaments of the myofibrils.
CC -!- DOMAIN: The rodlike tail sequence is highly repetitive, showing
CC cycles of a 28-residue repeat pattern composed of 4 heptapeptides,
CC characteristic for alpha-helical coiled coils.
CC -!- MISCELLANEOUS: Each myosin heavy chain can be split into 1 light
CC meromyosin (LMW) and 1 heavy meromyosin (HMM). It can later be
CC split further into 2 globular subfragments (S1) and 1 rod-shaped
CC subfragment (S2).
CC -!- SIMILARITY: Contains 1 myosin-like globular head domain.
```


Muss W., Kametner R., Klaussegger A., Huber A., Pohla-Gubo G.,
Wiche G., Uitto J., Hintner H.;
"A compound heterozygous one amino-acid insertion/nonsense mutation in
the plectin gene causes epidermolysis bullosa simplex with plectin
deficiency.";
Am. J. Pathol. 158:617-625(2001).
[5]
VARIANT EBS1 TRP-2110.
MEDLINE=21841370; PubMed=11851880;
Koss-Harnes D., Hoeyheim B., Anton-Lamprecht I., Gjesti A.,
Joergensen R.S., Jahnson F.L., Olaisen B., Wiche G.,
Gedde-Dahl T. Jr.;
"A site-specific plectin mutation causes dominant epidermolysis
bullosa simplex Ogna: two identical de novo mutations.";
J. Invest. Dermatol. 118:87-93(2002).
CC -!- FUNCTION: Interlinks intermediate filaments with microtubules and
microfilaments and anchors intermediate filaments to desmosomes or
hemidesmosomes. Could also bind muscle proteins such as actin to
membrane complexes in muscle. May be involved not only in the
crosslinking and stabilization of cytoskeletal intermediate
filaments network, but also in the regulation of their dynamics.
CC -!- SUBUNIT: Homodimer or homotetramer.
CC -!- ALTERNATIVE PRODUCTS:
CC Event-Alternative splicing; Named isoforms=3;
CC Name=1;
CC IsoId=Q15149-1; Sequence=Displayed;
CC Name=2;
CC IsoId=Q15149-2; Sequence=VSP_005030;
CC Name=3;
CC IsoId=Q15149-3; Sequence=VSP_005030, VSP_005031;
CC -!- TISSUE SPECIFICITY: Widely expressed with highest levels in
muscle, heart, placenta and spinal cord.
CC -!- DOMAIN: The N-terminus interacts with actin, the C-terminus with
vimentin, desmin, GFAP, cytokeratins, lamin B; whereas both the N-
and the C-terminus can bind integrin beta-4.
CC -!- PTM: PHOSPHORYLATED BY CDC2; REGULATES DISSOCIATION FROM
INTERMEDIATE FILAMENTS DURING MITOSIS (BY SIMILARITY).
CC -!- DISEASE: Defects in PLEC1 are the cause of epidermolysis bullosa
simplex with muscular dystrophy (MD-BBS) [MIM:226670]; an
autosomal recessive disorder characterized by epidermal blister
formation at the level of the hemidesmosome and associated with
late-onset muscular dystrophy.
CC -!- DISEASE: Defects in PLEC1 are the cause of epidermolysis bullosa
simplex 1 (EBS1) [MIM:131950]; also called epidermolysis bullosa
simplex Ogna type. EBS1 is an autosomal dominant form of
epidermolysis bullosa simplex differentiated from the more
generalized form of Koebner [MIM:131900] and the localized form of
Weber and Cockayne [MIM:131800] by the occurrence of skin
bruising.
CC -!- SIMILARITY: Contains 1 actin-binding domain.
CC -!- SIMILARITY: Contains 2 calponin-homology (CH) domains.
CC -!- SIMILARITY: Contains 33 plectin repeats.
CC -!- SIMILARITY: Contains 4 spectrin repeats.
CC -!- SIMILARITY: Belongs to the plectin or cytolinker family.

THIS SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
or send an email to license@isb-sib.ch).

CC EMBL; Z54367; CAA91196.1; -
CC EMBL; U53204; AAB05427.1; -
CC EMBL; U63610; AAB05428.1; -
CC EMBL; U63609; AAB05428.1; JOINED.
CC EMBL; X97053; CAA65765.1; -
CC PIR; C59404; A59404.
CC HSPG; Q01082; 1BRK.
CC Genew; HGNC:9069; PLEC1.
CC GK; Q15149; -
CC MIM; 601282; -

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Cricetinae;
OC Mesocricetus.
OX NCBI_TaxID=10036;
RN [1]
RN SEQUENCE FROM N.A.
RP STRAIN=F1B; TISSUE=Liver;
RX MEDLINE=95115033; PubMed=7815459;
RA Wang R., Sole M.J., Cukerman E., Liew C.-C.;
RT "Characterization and nucleotide sequence of the cardiac alpha-myosin
heavy chain gene from Syrian hamster."
RL J. Mol. Cell. Cardiol. 26:1155-1165(1994).
RN [2]
RP SEQUENCE OP 962-1935 FROM N.A.
RX MEDLINE=88247788; PubMed=3380703;
RA Jandreski M.A., Sole M.J., Liew C.-C.;
RT "Sequence of cDNA encoding the Syrian hamster cardiac beta-myosin
heavy chain."
RL Nucleic Acids Res. 16:4737-4737(1988).
CC -!- FUNCTION: Muscle contraction.
CC -!- SUBUNIT: Muscle myosin is a hexameric protein that consists of 2
heavy chain subunits (MHC), 2 alkali light chain subunits (MLC)
and 2 regulatory light chain subunits (MHC-2).
CC -!- SUBCELLULAR LOCATION: Thick filaments of the myofibrils.
CC -!- DOMAIN: The rodlike tail sequence is highly repetitive, showing
cycles of a 28-residue repeat pattern composed of 4 heptapeptides,
characteristic for alpha-helical coiled coils.
CC -!- MISCELLANEOUS: Each myosin heavy chain can be split into 1 light
meromyosin (LMM) and 1 heavy meromyosin (HMM). It can later be
split further into 2 globular subfragments (S1) and 1 rod-shaped
subfragment (S2).
CC -!- SIMILARITY: Contains 1 myosin-like globular head domain.
CC -!- SIMILARITY: Contains 1 IQ domain.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>
or send an email to license@isb-sib.ch).
DR EMBL; L12104; AAA62313.1; -
DR EMBL; X07273; CAA30256.1; -
DR F1R; I48153; 148153.
DR HSP; P08799; IMND.
DR InterPro: IPR000048; IQ region.
DR InterPro: IPR001609; myosin head.
DR InterPro: IPR004009; Myosin N.
DR InterPro: IPR002928; Myosin_tail.
DR Pfam; PF00612; IQ; 2.
DR Pfam; PF00663; myosin head; 1.
DR Pfam; PF02736; Myosin_N; 1.
DR Pfam; PF01576; Myosin_tail; 1.
DR PRINTS; PR00193; MYOSINHEAVY.
DR ProDom; PD000355; myosin_head; 1.
DR SMART; SM00015; IQ; 1.
DR SMART; SM00242; MYSC; 1.
DR PROSITE; PS0096; IQ; 1.
KW Myosin; Muscle protein; Coiled coil; Thick filament; Actin-binding;
KW ATP-binding; Myofibril; Multigene family; Calmodulin-binding.
FT DOMAIN 1 779
FT MYOSIN HEAD-LIKE.
FT DOMAIN 780 809
FT IQ.
FT DOMAIN 839 1934
FT COILED COIL (POTENTIAL).
FT NP_BIND 177 184
FT ATP.
FT ACTIN-BINDING.
FT DOMAIN 654 676
FT DOMAIN 756 770
FT METHYLATION (TRI-) (POTENTIAL).
FT MOD_RES 128 128
FT D -> E (IN REF. 2).
FT T -> TE (IN REF. 2).
FT CONFLICT 978 978
FT E -> Q (IN REF. 2).
FT CONFLICT 985 985
FT CONFLICT 1008 1014
FT DLAQED -> ALRAKT (IN REF. 2).
FT CONFLICT 1057 1057

FT CONFLICT 1060 1060 L -> V (IN REF. 2).
FT CONFLICT 1095 1095 D -> N (IN REF. 2).
FT CONFLICT 1217 1217 E -> D (IN REF. 2).
FT CONFLICT 1271 1271 D -> N (IN REF. 2).
FT CONFLICT 1327 1327 T -> A (IN REF. 2).
FT CONFLICT 1358 1358 C -> R (IN REF. 2).
FT CONFLICT 1504 1504 L -> V (IN REF. 2).
FT CONFLICT 1537 1537 M -> L (IN REF. 2).
FT CONFLICT 1556 1556 N -> K (IN REF. 2).
SQ SEQUENCE 1934 AA; 222928 MW; PDBAC58310B0B57D CRC64;
Query Match 8.8%; Score 237; DB 1; Length 1934;
Best Local Similarity 23.3%; Pred. No. 0.015;
Matches 129; Conservative 88; Mismatches 208; Indels 128; Gaps 20;
QY 12 EGQRRPSQAAPAVEAEGPGSSQAPRKPEGAQARTAQSGALRDVSELSQLSDIILSTYC 71
DB 1034 EGSLEQEKVMDLE-----RAKKLEGDLKLTQBS--IMDLEND-KQOLEKLLKKD 1083
QY 72 VDNQGGPGEDGAGCPAPEDA-----EKSTTYVAR-----NGEPETTPVYGEKEP 119
DB 1084 FELN-----ALNARIEDEQALGSQLQKKLQELQARIEELBELEAEARTAKVEKLR 1135
QY 120 SKGDPNTEIROS-DEVG-----DRDHRPPOKKKAKGLGKEITILLMOTLNTLSTPEE 171
DB 1136 SLSRELEISELERAGGATSVQIENKKREAEFFQWRDLLEATL-----QHEA 1186
QY 172 KLAALCKKY-----ABLEEHRSQKMKLLQKKSQVQKQHLRGHSHKAVLARSKLES 227
DB 1187 TAAALRKHADSVAELEGQIDNLRVQKLEKESEKSEFLEDDVTSMNQEIIRAKANLEK 1246
QY 228 LCR-----ELQRHN-----RSLKE----- 241
DB 1247 MCTLEQDMEHRSKAEETORSVNDLTSQRAKQOTENGELSRLQDEKALISQLTRGKLT 1306
QY 242 -----EGVQRRAREEEXRKEVTSH-FQWTLNDIQLQMEQHNERNK-----LRQENME 288
DB 1307 YTQQLDLKLEEVKAKNTLAHALQSAHSDCDLLREQVEEETAKAELQCVLSKANSE 1366
QY 289 LAELKKLIEQYELREBHIDKVPKDLQOOLVDKLOQAOE-----MLKAEERHOR 341
DB 1367 VAQWRTYETDAIQTBELSEA--KKLAQRLQDA--BEAVEAVNAKSSLEKTKHRLQN 1422
QY 342 EKDFLLKAVESQRCBIMKQOETHLKQALALYTERKEEFQNTLSKSEVFTTPKQENKEK 401
DB 1423 EIEDLWVDVRSNAAAALDKQKQNPDKILAENKQKYEESQSELESQKARSLSLTEL-- 1480
QY 402 MTKIKKLEKETTYRGRWSSNKALLEMAEKTVRDKLEGLQVQIKORLEKLCALQTE 461
DB 1481 --FKLKNAYEESLEHLETFKRENKQLEISDLT---EQLGSTGKSIHLEKIRKQLEAE 1535
QY 462 RNDLKKRVQDLQA 474
DB 1536 KNELOSALAEABA 1548
RESULT 25
MYHD HUMAN
ID MYHD HUMAN STANDARD; PRT; 1938 AA.
AC Q9UKX3; OS252;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 15-MAR-2004 (Rel. 43, Last annotation update)
DE Myosin heavy chain, skeletal muscle, extraocular (MyHC-eo).
GN MYH13.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RN SEQUENCE FROM N.A.
RP TISSUE=Extraocular muscle;
RC TISSUE=Extraocular muscle;
RC MEDLINE=99318669; PubMed=10388558;
RX

FT	657	679	ACTIN-BINDING.
FT	759	773	ACTIN-BINDING.
FT	129	129	METHYLATION (TRI-) (POTENTIAL) .
FT	88	88	Q -> E (IN REF. 1).
FT	574	574	Q -> P (IN REF. 1).
FT	608	608	A -> G (IN REF. 1).
FT	744	744	T -> A (IN REF. 1).
FT	790	790	M -> I (IN REF. 1).
FT	1014	1014	V -> A (IN REF. 1).
FT	1021	1021	S -> T (IN REF. 1).
FT	1101	1101	A -> V (IN REF. 1).
FT	1290	1290	A -> S (IN REF. 1).
FT	1373	1373	W -> C (IN REF. 1).
FT	1533	1533	K -> N (IN REF. 1).
FT	1540	1540	L -> M (IN REF. 1).
FT	1578	1578	KL -> NV (IN REF. 5).
FT	1705	1705	EQ -> DR (IN REF. 1).
FT	1733	1733	E -> D (IN REF. 1).
FT	1734	1734	A -> S (IN REF. 2).
FT	1737	1737	T -> S (IN REF. 1).
FT	1763	1763	D -> H (IN REF. 1).
FT	1788	1788	M -> I (IN REF. 3).
FT	1871	1871	D -> N (IN REF. 5).
FT	1882	1882	R -> G (IN REF. 5).
FT	1890	1890	Q -> R (IN REF. 5).
FT	1933	1933	MISSING (IN REF. 5).
FT	1939	223689	MM; ECB87E7CEB768B6F CRC64;
FT	1939	AA;	
Query Match	8.8%;	Score 236.5;	DB 1; Length 1939;
Best Local Similarity	24.4%;	Pred. No. 0.016;	
Matches 146;	Conservative 81;	Mismatches 196;	Indels 175; Gaps 25;
QY	12	EGAQEPSSQAARAVEABGPGSSQAQPRKPEGAQAARTAQSGALPDVSEELSRQLEDILSTYC 71	
Db	1037	EGSLSEKKEKVRMDLSE-----RAKKLEGGDLKTQES--IMDLND-KLQEEELKKKE 1086	
QY	72	VNNNGGPGEDGAQGPAPEDA-----EKSRTVVAR-----NGEPTPTVVVGEKRP 119	
Db	1087	FDINQ-----QNSKIDEQALALQLOKLEKQARLEEEELAEARTAPAKVEKL 1138	
QY	120	SKGDPNTEIROS-DEVG-----DRHRRPOEKKKAGKIGKEITLLMOTLNTLSTPEE 171	
Db	1139	SDLSRELEISERLEBAGGATSVQIEMNKKRAEFQKMRDLDEEATL-----OHEA 1189	
QY	172	KLALCKKY-----AEILLEHRNSQOMKLQKOSQLOVEKDLHARGEHSKAVLARSKLES 227	
Db	1190	TAAALRKGHADSVAEIGEQIDNLRQVKQLEKESEFKLELDDVTSNMEOIIXAKANLEK 1249	
QY	228	LCH-----ELQRNRSIKESGVQAR----- 248	
Db	1250	VSTLEDQANEYRVKLEEAQRSLNDPTTQAKLOTENGELARLEKEALISQTRGKLS 1309	
QY	249	-----BEERKREVTSH-FOVTLNDIOLQMEQHNR-----NSK 281	
Db	1310	YTOQMEDKALQLEEEGKAKNALAHQARSARHDCDLLREQVYBEETAEKAEQLQRVLSKANSE 1369	
QY	282	LR-----QENNELAERLKKLIEQVELREEHIDKY-----PMKH-----DLQ 318	
Db	1370	VAGWRKYETDALQRTTEELLEAAKCKLAQRIQDAEABEAVNAKSCSSILEKTKHRLQNEIED 1429	
QY	319	QLVDAKLQQAQEMLKABERHQBREKDFLKEAVESQRMCELMKXQOETHLKQOLALYTEK 378	
Db	1430	LMTDVERSNA-----AAAALDKQRNFKILAE--WKQKYBESQSELESQKARSLSLELF 1484	
QY	379	EETONTLSKSESVPTFKQEMERQTKI-----KKLEKETTMWR 417	
Db	1485	KUNAYBESLEHLETFKRNKNLQBEISDLTQOLGGGKNVHELEKVRQOLEVEXKLELQ 1543	
QY	418	SRWESSNKALLEWAEKTVVDKBELEGQVKIQLEKLC-----RALQTERNDLAKRVQD 471	
Db	1544	SALSEA-EASLEHEEGKILR-AQLEBFNQIKAEIERKLAEDKBEAKRN--HORVVD 1597	

RA Kutateladze T.G., Ogburn K.D., Watson W.T., de Beer T., Emr S.D.,
RA Burd C.G., Overduin M.;
RT "Phosphatidylinositol 3-phosphate recognition by the FYVE domain.";
RL Mol. Cell 3:805-811(1999).
[7]
RN MUTAGENESIS OF TRP-1349; CYS-1358; PHE-1365; ARG-1370; ARG-1371;
RP HIS-1372; HIS-1373; CYS-1374; ARG-1375; CYS-1377; GLY-1378; CYS-1385;
RP ARG-1400 AND CYS-1405, SUBCELLULAR LOCATION, AND INTERACTION WITH
RP PHOSPHATIDYLINOSITOL 3-PHOSPHATE.
RX MEDLINE=20387352; PubMed=10807926;
RA Gaullier J.-M., Roemling E., Gillooly D.J., Stenmark H.;
RT "Interaction of the EEA1 FYVE finger with phosphatidylinositol
RT 3-phosphate and early endosomes. Role of conserved residues.";
RL J. Biol. Chem. 275:24595-24600(2000).
[8]
RN INTERACTION WITH RAB22A.
RX MEDLINE=21859373; PubMed=11870209;
RA Kauppi M., Simonsen A., Bremnes B., Vieira A., Callaghan J.M.,
RA Stenmark H., Olkkonen V.M.;
RT "The small GTPase Rab22 interacts with EEA1 and controls endosomal
RT membrane trafficking.";
RL J. Cell Sci. 115:899-911(2002).
[9]
RN MUTAGENESIS OF GLU-39; PHE-41; ILE-42; PRO-44; MET-47 AND TYR-60,
RP HOMODIMERIZATION, AND INTERACTION WITH RAB5C.
RX MEDLINE=22495593; PubMed=12493736;
RA Merithew E., Stone C., Bathiraj S., Lambright D.G.;
RT "Determinants of Rabs' interaction with the N terminus of early
RT endosome antigen 1.";
RL J. Biol. Chem. 278:8494-8500(2003).
[10]
RN X-RAY CRYSTALLOGRAPHY (2.2 ANGSTROMS) OF 1289-1411 IN COMPLEX WITH
RP PHOSPHATIDYLINOSITOL 3-PHOSPHATE, AND HOMODIMERIZATION.
RX MEDLINE=21617582; PubMed=11741531;
RA Dumas J.J., Merithew E., Sudharshan E., Rajamani D., Hayes S.,
RA Lawe D., Corvera S., Lambright D.G.;
RT "Multivalent endosome targeting by homodimeric EEA1.";
RL Mol. Cell 8:947-958(2001).
[11]
RN STRUCTURE BY NMR OF 1346-1410 ALONE AND IN COMPLEX WITH
RP PHOSPHATIDYLINOSITOL 3-PHOSPHATE.
RX MEDLINE=21143489; PubMed=11230696;
RA Kutateladze T.G., Overduin M.;
RT "Structural mechanism of endosome docking by the FYVE domain.";
RL Science 291:1793-1796(2001).
CC -I- FUNCTION: Binds phospholipid vesicles containing
CC phosphatidylinositol 3-phosphate and participates in endosomal
CC trafficking.
CC -I- SUBUNIT: Homodimer. Binds STX6. Binds RAB5A, RAB5B, RAB5C and
CC RAB22A that have been activated by GTP-binding.
CC -I- SUBCELLULAR LOCATION: Cytoplasmic and peripheral membrane protein
CC associated with early endosomes.
CC -I- DOMAIN: The FYVE-type zinc finger domain mediates interactions
CC with phosphatidylinositol 3-phosphate.
CC -I- DISEASE: Antibodies against EEA1 are found in sera from patients
CC with subacute cutaneous lupus erythematosus and other autoimmune
CC diseases.
CC -I- SIMILARITY: Contains 1 C2H2-type zinc finger.
CC -I- SIMILARITY: Contains 1 FYVE-type zinc finger.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).

CC EMBL: L40157; AAA79121.1; -;
DR EMBL: X78998; CAAS5632.1; -;
DR PIR: A57013; A57013.
DR PDB: 1HYI; 14-MAR-01.
DR PDB: 1HYJ; 14-MAR-01.

DR PDB: 1JOC; 28-DEC-01.
DR Genew; HGNC:3185; EEA1.
DR MIM; 605070; -.
DR GO; GO:0005737; C:cytoplasm; TAS.
DR GO; GO:0005769; C:early endosome; TAS.
DR GO; GO:0005624; C:membrane fraction; TAS.
DR GO; GO:0005545; F:phosphatidylinositol binding; TAS.
DR GO; GO:0006899; P:nonspecific vesicle transport; TAS.
DR InterPro; IPR007087; Znf_C2H2.
DR InterPro; IPR000306; Znf_FYVE.
DR Pfam; PF01363; FYVE; 1.
DR Pfam; PF00096; zf-C2H2; 1.
DR SMART; SM00064; FYVE; 1.
DR SMART; SM00355; Znf_C2H2; 1.
DR PROSITE; PS0178; ZF_FYVE; 1.
DR PROSITE; PS00028; ZINC_FINGER_C2H2_2; 1.
DR PROSITE; PS0157; ZINC_FINGER_C2H2_2; 1.
KW Antigen; Metal-binding; Zinc; Coiled coil; Zinc-finger; 3D-structure.
FT DOMAIN 74 1348
FT COILED COIL (POTENTIAL).
FT DOMAIN 97 758
FT GLN/GLU/LYS-RICH.
FT DOMAIN 1093 1231
FT GLN/GLU/LYS-RICH.
FT ZN_FING 41 64
FT C2H2-TYPE.
FT ZN_FING 1352 1410
FT FYVE-TYPE.
FT MUTAGEN 39 39
FT E->A: STRONGLY REDUCES INTERACTION WITH
FT RAB5C.
FT MUTAGEN 41 41
FT F->A: STRONGLY REDUCES INTERACTION WITH
FT RAB5C.
FT MUTAGEN 42 42
FT I->A: STRONGLY REDUCES INTERACTION WITH
FT RAB5C.
FT MUTAGEN 44 44
FT P->A: STRONGLY REDUCES INTERACTION WITH
FT RAB5C.
FT MUTAGEN 47 47
FT M->A: STRONGLY REDUCES INTERACTION WITH
FT RAB5C.
FT MUTAGEN 60 60
FT Y->A: STRONGLY REDUCES INTERACTION WITH
FT RAB5C.
FT W->A: REDUCES PHOSPHATIDYLINOSITOL 3-
FT PHOSPHATE BINDING AND ENDOSOMAL LOCATION.
FT D->V: REDUCES PHOSPHATIDYLINOSITOL 3-
FT PHOSPHATE BINDING AND ENDOSOMAL LOCATION.
FT N->D: REDUCES PHOSPHATIDYLINOSITOL 3-
FT PHOSPHATE BINDING AND ENDOSOMAL LOCATION.
FT C->S: ABOLISHES PHOSPHATIDYLINOSITOL 3-
FT PHOSPHATE BINDING AND ENDOSOMAL LOCATION.
FT P->A: STRONGLY REDUCES
FT PHOSPHATIDYLINOSITOL 3-PHOSPHATE BINDING
FT AND ENDOSOMAL LOCATION.
FT VT->EE,GG: ABOLISHES
FT PHOSPHATIDYLINOSITOL 3-PHOSPHATE BINDING
FT AND ENDOSOMAL LOCATION.
FT R->A: ABOLISHES ENDOSOMAL LOCATION.
FT R->A: ABOLISHES PHOSPHATIDYLINOSITOL 3-
FT PHOSPHATE BINDING AND ENDOSOMAL LOCATION.
FT H->A: ABOLISHES PHOSPHATIDYLINOSITOL 3-
FT PHOSPHATE BINDING AND ENDOSOMAL LOCATION.
FT C->A: ABOLISHES PHOSPHATIDYLINOSITOL 3-
FT PHOSPHATE BINDING AND ENDOSOMAL LOCATION.
FT R->G: ABOLISHES PHOSPHATIDYLINOSITOL 3-
FT PHOSPHATE BINDING AND ENDOSOMAL LOCATION.
FT C->A: ABOLISHES PHOSPHATIDYLINOSITOL 3-
FT PHOSPHATE BINDING AND ENDOSOMAL LOCATION.
FT G->A: ABOLISHES PHOSPHATIDYLINOSITOL 3-
FT PHOSPHATE BINDING AND ENDOSOMAL LOCATION.
FT C->A: ABOLISHES PHOSPHATIDYLINOSITOL 3-
FT PHOSPHATE BINDING AND ENDOSOMAL LOCATION.
FT R->G: STRONGLY REDUCES
FT PHOSPHATIDYLINOSITOL 3-PHOSPHATE BINDING
FT AND ABOLISHES ENDOSOMAL LOCATION.
FT C->S: ABOLISHES PHOSPHATIDYLINOSITOL 3-
FT PHOSPHATE BINDING AND ENDOSOMAL LOCATION.
FT C->S (IN REF. 1).
FT LQ -> FE (IN REF. 1).

```

FT CONFLICT 277 277 A -> S (IN REF. 1).
FT CONFLICT 284 284 A -> R (IN REF. 1).
FT CONFLICT 520 520 D -> E (IN REF. 1).
FT CONFLICT 575 576 EQ -> DE (IN REF. 1).

Query Match 8.7%; Score 2367; DB 1; Length 1411;
Best Local Similarity 23.1%; Pred. No. 0.012;
Matches 129; Conservative 90; Mismatches 200; Indels 140; Gaps 22;

QY 6 QPEAGPGAGRRPQCAAPAEAGGSSQAPRPGAGCAARTAGSALRDVSE-----ELSR 61
D 121 QOEAKPDLVTDSSAELQSLSE-----QQLEEAQTFENFIKQKDLFEQKAAQALAT 170
QY 62 QLEDILSTYCVNDNQGSGEDGAGQGEPAEPEDAKSRITYVARNGSEPPTVPVYGEKPSK 121
D 171 EIADIKSKY-----DEERSLRAARQKVTIRTEELNKAT-VIQDLKTELL 215
QY 122 GDNPTERI-----RQSEVGDGRDHRPQE-----KK 147
D 216 QRPGLDVAVLKELVQVOTLMDNNTLERERESEKLDCKEKLQSQYASSEATISQLRSE 275
QY 148 KAKGLGKEITLMTLNTLSTPEEKJAA-----LCKK---VAELLEEHRNSQKQMKL 196
D 276 LAKG-POEVAVYVQELQKLSVNELOFQKNTLNTLNLKKEQDTYKLEEKHNEESVSKN 334
QY 197 L-----QKQSQVQEKDHLRGEH---SKAVLARSKLSESLCRELQHRNRSKKE 242
D 335 IQATLHQKDLQDQQLQSLASSETSLHRIHVSEKGEATQKLEELSEVETKYOHLKAE 394
QY 243 GVQARABEEERKVEVSHFQVTLNDIQLOMEQHNRSLKPOENMELARLKLIEQVEL 302
D 395 FKQLOQOQREKQ-----HGLQLOQE-INQLSHKLQETTERQGEAHRKQKQRL 443
QY 303 REEHIDKVFHKDLQQLQVLD--AKLQQAQEMLKE-----AEERHQREKDFLLKAEVSR 355
D 444 SSE-----KLMDKEQVADILQLKSLREELQKKEKVTNSTELQHLQDK---TKQHQEQ 494
QY 356 MCE-----LMQOETHLQQLALYTEKEEFQN---TLKSEVFTTFKQEMKWTYKIK 407
D 495 ALQOQSTAKLREAOQDLQVLRQIGDKDQKQNTLEALLOKSKENISLLEKREEDLYAKIQ 554
QY 408 KLEKETTIVRSWESSNKALLEMAEKTVRDKE-----LEGLOVKIQRLKELCRA-- 457
D 555 AGEGETAVL-NQLOEKNTLQEQVTLQTEKLKNQSESHKQAEHLNDQVQKQKHLRAAQ 613
QY 458 -----LOTERNDLNKRQVD 471
D 614 DVLVLSLETSLVNEINQSLNE 632

RESULT 29
MYH7_HUMAN
ID MYH7_HUMAN STANDARD; PRT; 1935 AA.
AC P12883; Q14904; Q16579;
DT 01-OCT-1989 (Rel. 12, Created)
DT 01-JUN-1994 (Rel. 29, Last sequence update)
DT 15-MAR-2004 (Rel. 43, Last annotation update)
DE Myosin heavy chain, cardiac muscle beta isoform (MYHC-beta).
GN MYH7 OR MYSCB.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]_
RP SEQUENCE FROM N.A.
RX MEDLINE=91065334; PubMed=2249844;
RA Jaenicke T., Dieckrich K.W., Haas W., Schleich J., Lichter P.,
RA Pfordt M., Bach A., Vosberg H.P.;
RT "The complete sequence of the human beta-myosin heavy chain gene and
RT a comparative analysis of its product.";
RL Genomics 8:194-206(1990).
RN [2]
RP SEQUENCE FROM N.A.

```

```

RX MEDLINE=90301496; PubMed=2362820;
RA Liew C.-C., Sole M.J., Yamauchi-Takahara K., Kellam B.,
RA Anderson D.H., Lin L., Liew J.,
RT "Complete sequence and organization of the human cardiac beta-myosin
RT heavy chain gene.";
RL Nucleic Acids Res. 18:3647-3651(1990).
RN [3]
RP SEQUENCE OF 1-115 FROM N.A.
RX MEDLINE=89264452; PubMed=2726733;
RA Yamauchi-Takahara K., Sole M.J., Liew J., Ing D., Liew C.-C.;
RT "Characterization of human cardiac myosin heavy chain genes.";
RL Proc. Natl. Acad. Sci. U.S.A. 86:3504-3508(1989).
RN [4]
RP ERRATUM.
RA Yamauchi-Takahara K., Sole M.J., Liew J., Ing D., Liew C.-C.;
RL Proc. Natl. Acad. Sci. U.S.A. 86:7416-7417(1989).
RN [5]
RP SEQUENCE OF 1310-1935 FROM N.A.
RX MEDLINE=86176778; PubMed=2421254;
RA Saez L., Leinwand L.A.;
RT "Characterization of diverse forms of myosin heavy chain expressed in
RT adult human skeletal muscle.";
RL Nucleic Acids Res. 14:2951-2969(1986).
RN [6]
RP REVISIONS.
RA Leinwand L.A.;
RL Submitted (MAR-1988) to the EMBL/GenBank/DBJ databases.
RN [7]
RP SEQUENCE OF 1410-1935 FROM N.A.
RX MEDLINE=88299163; PubMed=2969919;
RA Kurabayashi M., Tsuchimochi H., Komuro I., Takaku F., Yazaki Y.;
RT "Molecular cloning and characterization of human cardiac alpha- and
RT beta-form myosin heavy chain complementary DNA clones. Regulation of
RT expression during development and pressure overload in human
RT atrium.";
RL J. Clin. Invest. 82:524-531(1988).
RN [8]
RP SEQUENCE OF 785-1935 FROM N.A.
RC TISSUE=Skeletal muscle;
RX MEDLINE=90235882; PubMed=1691980;
RA Bober E., Buchberger-Seidl A., Braun T., Singh S., Goedde H.W.,
RA Arnold H.H.;
RT "Identification of three developmentally controlled isoforms of human
RT myosin heavy chains.";
RL Eur. J. Biochem. 189:55-65(1990).
RN [9]
RP SEQUENCE OF 1393-1935 FROM N.A.
RX MEDLINE=87192738; PubMed=3032769;
RA Jandreski M.A., Liew C.-C.;
RT "Construction of a human ventricular cDNA library and
RT characterization of a beta myosin heavy chain cDNA clone.";
RL Hum. Genet. 76:47-53(1987).
RN [10]
RP REVIEW ON VARIANTS.
RX MEDLINE=96039076; PubMed=8533830;
RA Arai S., Matsuo K., Hirayama K., Sukurai H., Tamura M., Ozawa T.,
RA Kimura M., Imamura S.-I., Furuta Y., Joh-O K., Kawana M., Takao A.,
RA Hosoda S., Momma K.;
RT "Missense mutation of the beta-cardiac myosin heavy-chain gene in
RT hypertrophic cardiomyopathy.";
RL Am. J. Med. Genet. 58:267-276(1995).
RN [11]
RP VARIANTS CMH1 GLU-256 AND ARG-741.
RX MEDLINE=93248216; PubMed=8483915;
RA Fananapazir L., Dalakas M.C., Cyran P., Cohn G., Epstein N.D.;
RT "Missense mutations in the beta-myosin heavy-chain gene cause central
RT core disease in hypertrophic cardiomyopathy.";
RL Proc. Natl. Acad. Sci. U.S.A. 90:3993-3997(1993).
RN [12]
RP VARIANT CMH1 GLN-403.
RX MEDLINE=90367131; PubMed=1975517;
RA Geisterfer-Lowrance A.A.T., Kass S., Tanigawa G., Vosberg H.-P.,
RA McKenna W., Seidman C.B., Seidman J.G.;

```

"A molecular basis for familial hypertrophic cardiomyopathy: a beta cardiac myosin heavy chain gene missense mutation.";
 Cell 62:999-1006(1990).
 [13]
 RP VARIANTS CMH1 GLN-249; GLN-403; CYS-453 AND MET-606.
 RX MEDLINE=92204193; PubMed=1552912;
 RA Watkins H., Rosenzweig A., Hwang D.S., Levi T., McKenna W., Seidman C.E., Seidman J.G.;
 RA "Characteristics and prognostic implications of myosin missense mutations in familial hypertrophic cardiomyopathy.";
 RL New Engl. J. Med. 326:1108-1114(1992).
 [14]
 RP VARIANTS CMH1 GLN-403; CYS-453; ARG-584 AND MET-606.
 RX MEDLINE=94070863; PubMed=8250038;
 RA Watkins H., Thierfelder L., Anan R., Jarcho J., Matsumori A., McKenna W., Seidman J.G., Seidman C.E.;
 RA "Independent origin of identical beta cardiac myosin heavy-chain mutations in hypertrophic cardiomyopathy.";
 RL Am. J. Hum. Genet. 53:1180-1185(1993).
 [15]
 RP VARIANTS CMH1 GLN-403 AND VAL-908.
 RX MEDLINE=92346810; PubMed=1638703;
 RA Epstein N.D., Cohn G.M., Cyran F., Pananapazir L.;
 RA "Differences in clinical expression of hypertrophic cardiomyopathy associated with two distinct mutations in the beta-myosin heavy chain gene. A 908Leu-->Val mutation and a 403Arg-->Gln mutation.";
 RL Circulation 86:345-352(1992).
 [16]
 RP VARIANTS CMH1 LEU-403 AND TRP-403.
 RX MEDLINE=94075629; PubMed=8254035;
 RA Dausse E., Komajda M., Feller L., Dubourg O., Dufour C., Carrier L., Wisniewsky K., Bercovic J., Hengstenberg C., Al-Mahdawi S.;
 RA "Familial hypertrophic cardiomyopathy. Microsatellite haplotyping and identification of a hot spot for mutations in the beta-myosin heavy chain gene.";
 RL J. Clin. Invest. 92:2807-2813(1993).
 [17]
 RP VARIANT CMH1 TRP-403.
 RX MEDLINE=94093568; PubMed=8268932;
 RA Moolman J.C., Brink P.A., Corfield V.A.;
 RA "Identification of a new missense mutation at Arg403, a CpG mutation hotspot, in exon 13 of the beta-myosin heavy chain gene in hypertrophic cardiomyopathy.";
 RL Hum. Mol. Genet. 2:1731-1732(1993).
 [18]
 RP VARIANT CMH1 ASN-615.
 RX MEDLINE=93038688; PubMed=1417858;
 RA Nishi H., Kimura A., Harada H., Toshima H., Sasazuki T.;
 RA "Novel missense mutation in cardiac beta myosin heavy chain gene found in a Japanese patient with hypertrophic cardiomyopathy.";
 RL Biochem. Biophys. Res. Commun. 188:379-387(1992).
 [19]
 RP VARIANT CMH1 GLY-778.
 RX MEDLINE=93343938; PubMed=8343162;
 RA Harada H., Kimura A., Nishi H., Sasazuki T., Toshima H.;
 RA "A missense mutation of cardiac beta-myosin heavy chain gene linked to familial hypertrophic cardiomyopathy in affected Japanese families.";
 RL Biochem. Biophys. Res. Commun. 194:791-798(1993).
 [20]
 RP VARIANT CMH1 VAL-908.
 RX MEDLINE=93168485; PubMed=8435239;
 RA Al-Mahdawi S., Chamberlain S., Cleland J., Nihoyannopoulos P., Gilligan D., French J., Choudhury L., Williamson R., Oakley C.;
 RA "Identification of a mutation in the beta cardiac myosin heavy chain gene in a family with hypertrophic cardiomyopathy.";
 RL Br. Heart J. 69:136-141(1993).
 [21]
 RP VARIANT CMH1 TRP-719.
 RX MEDLINE=95179132; PubMed=7874131;
 RA Greve G., Bachinski L., Friedman D.L., Czernuszewicz G., Anan R., Towbin J., Seidman C.E., Roberts R.;
 RA "Isolation of a de novo mutant myocardial beta MHC protein in a

pedigree with hypertrophic cardiomyopathy.";
 Hum. Mol. Genet. 3:2073-2075(1994).
 [22]
 RP VARIANTS CMH1 CYS-513; ARG-716 AND TRP-719.
 RX MEDLINE=94110336; PubMed=8282798;
 RA Anan R., Greve G., Thierfelder L., Watkins H., McKenna W., Solomon S., Vecchio C., Shono H., Nakao S., Mares A. Jr., Towbin J.A., Spirito P., Roberts R., Seidman J.G., Seidman C.E.;
 RA "Prognostic implications of novel beta cardiac myosin heavy chain gene mutations that cause familial hypertrophic cardiomyopathy.";
 RL J. Clin. Invest. 93:280-285(1994).
 [23]
 RP VARIANT CMH1 THR-797.
 RX MEDLINE=96047159; PubMed=7501410;
 RA Moolman J.C., Brink P.A., Corfield V.A.;
 RA "Identification of a novel Ala797Thr mutation in exon 21 of the beta-myosin heavy chain gene in hypertrophic cardiomyopathy.";
 RL Hum. Mutat. 6:197-198(1995).
 [24]
 RP VARIANT CMH1 CYS-453.
 RX MEDLINE=96209901; PubMed=8655135;
 RA Ko Y.-L., Chen J.-J., Tang T.-K., Cheng J.-J., Lin S.-Y., Liou Y.-C., Kuan P., Wu C.-W., Lien W.-P., Liew C.-C.;
 RA "Malignant familial hypertrophic cardiomyopathy in a family with a 453Arg-->Cys mutation in the beta-myosin heavy chain gene: coexistence of sudden death and end-stage heart failure.";
 RL Hum. Genet. 97:585-590(1996).
 [25]
 RP VARIANTS CMH1 THR-349 AND TRP-719.
 RX MEDLINE=98204402; PubMed=9544842;
 RA Jeschke B., Uhl K., Weist B., Schroder D., Meitinger T., Dohlemann C., Vosberg H.-P.;
 RA "A high risk phenotype of hypertrophic cardiomyopathy associated with Query Match 8.7%; Score 234; DB 1; Length 1935;
 Best Local Similarity 22.4%; Pred. No. 0.019;
 Matches 127; Conservative 100; Mismatches 177; Indels 162; Gaps 25;
 QY 8 EAGPEGAQERPSQAAPAV-----EAGPGSSQAPRKEGAQARTAQSCALRDVSEELS 60
 DB 1363 KANSEVAQWRKYVETDAIORTBELEAKKLAQAEAEAVYNAKCSLEKTKHRLQ 1422
 QY 61 RQEDILSTYCVDDNQGPGEDGAGCPAPEDAKSRITYVARNGEPEPTFVYGEKES 120
 DB 1423 NEIEDLM-----VDVERSNAAL-----DKQR 1447
 QY 121 KGPENTEROSDEVDGRHRRPOEKKAKGLGKITILM-----OTLNTLSTPEKLAAL 176
 DB 1448 NFDKILAEWKQYERSQSELESSQ--KEARSLSLTKLKNAYESLEHLETFKRENKL 1505
 QY 177 CKYAEALLLEHRNSOKMKLQKQSQVQEKDHLRG-----EHSKAVLARSKLE- 226
 DB 1506 QEEISDLTEQLGSSGKTIHELEKVRKQAEKQEWELQSALEAEASLEHEEGKILRAQLBP 1565
 QY 227 -----SICRELQRHNSL-----KEEG----- 243
 DB 1566 NQKAFIEKLAEKDEMEQAKRNHLRVVDSLSQTDATRRNRNEALRVKKGEGDLNEM 1625
 QY 244 -VQRA---REBEKEKVTSHFQVTLNDIQLQMEQHNRNSKLRQENLEAER----- 292
 DB 1626 EIQLSHANEMAAEAQVKVS-LQSLKQDTQIQDDAVRANDDLK-ENIAIVERRNNLLQA 1683
 QY 293 -----LKKLIEQYE-----LREEHIDKVFKEKDI-----QQQLVDAKLAQAEMLKE 334
 DB 1684 ELEEELRAVVEQTERSRKLABQSLIETSERVQLHSHQNTSLNQKQKQDADLSQLOTEVEE 1743
 QY 335 A--EERHOREKDFLLKEAVESQRM--ELMKQOET--HLKQOLALYTEFEFQNTLSKS 388
 DB 1744 AVQECENAEK---AKKAITDAAMWAELKKEQDTSAHLERMKNQMEQITIKDQHLRDEA 1800
 QY 389 SEV-FTTFQEMERKTKIKKLEKETTVYRSRWSSNKALLEMAEKTVDKLE-LEGLOV 446
 DB 1801 EQIALGKGKKQLQKLEARVRELENE-----LE-AEQK--RNAESVKGMRK 1842

Qy 447 KIQRLKLCALQTERNDLNKRVQDL 472
Db 1843 SERRIKELTYQTEBDRKNL-LRLQDL 1867

RESULT 30

MYH7_PIG STANDARD; PRT; 1935 AA.
AC P79293;
DT 15-DEC-1998 (Rel. 37, Created)
DT 15-DEC-1998 (Rel. 37, Last sequence update)
DT 15-MAR-2004 (Rel. 43, Last annotation update)
DE Myosin heavy chain, cardiac muscle beta isoform (MyHC-beta).
GN MYH7.
OS Sus scrofa (Pig).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.
OX NCBI_TaxID=9823;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Domesica;
RA Ko Y.L.;
RL Submitted (OCT-1996) to the EMBL/GenBank/DBSJ databases.
CC -!- FUNCTION: Muscle contraction.
CC -!- SURUNIT: Muscle myosin is a hexameric protein that consists of 2 heavy chain subunits (MHC), 2 alkali light chain subunits (MLC) and 2 regulatory light chain subunits (MLC-2).
CC -!- SUBCELLULAR LOCATION: Thick filaments of the myofibrils.
CC -!- DOMAIN: The rodlike tail sequence is highly repetitive, showing cycles of a 28-residue repeat pattern composed of 4 heptapeptides, characteristic for alpha-helical coiled coils.
CC -!- MISCELLANEOUS: Each myosin heavy chain can be split into 1 light meromyosin (LMW) and 1 heavy meromyosin (HMM). It can later be split further into 2 globular subfragments (S1) and 1 rod-shaped subfragment (S2).
CC -!- MISCELLANEOUS: The cardiac alpha isoform is a 'fast' ATPase myosin, while the beta isoform is a 'slow' ATPase.
CC -!- SIMILARITY: Contains 1 myosin-like globular head domain.
CC -!- SIMILARITY: Contains 1 IQ domain.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See [http://www.isb-sib.ch/](http://www.isb-sib.ch/announce/) or send an email to license@isb-sib.ch).
CC
CC EMBL; U75316; AAS37320.1; -
CC PIR; A59286; A59286.
CC HSP; P08799; 1MND.
CC InterPro; IPR000048; IQ region.
CC InterPro; IPR01609; myosin head.
CC InterPro; IPR004009; myosin N.
CC InterPro; IPR002828; myosin_tail.
CC Pfam; PF00612; IQ; 2.
CC Pfam; PF00063; myosin head; 1.
CC Pfam; PF02736; Myosin N; 1.
CC Pfam; PF01576; Myosin tail; 1.
CC PRINTS; PR00193; MYOSINHEAVY.
CC ProDom; PD000355; myosin_head; 1.
CC SMART; SM00015; IQ; 1.
CC SMART; SM00242; MSc; 1.
CC PROSITE; PS50096; IQ; 1.
CC Myosin; Muscle protein; Coiled coil; Thick filament; Actin-binding;
KW ATP-binding; Methylation; Multigene family; Calmodulin-binding.
FT DOMAIN 1 780 MYOSIN HEAD-LIKE.
FT DOMAIN 781 810 IQ.
FT DOMAIN 839 1925 COILED COIL (POTENTIAL).
FT NP_BIND 178 185 ATP (POTENTIAL).
FT DOMAIN 655 677 ACTIN-BINDING.
FT DOMAIN 757 771 ACTIN-BINDING.

FT MOD_RES 129 129 METHYLATION (TRI-) (POTENTIAL).
SQ SEQUENCE 1935 AA; 223110 MW; 149CDBF910DBB08 CRC64;
Query Match 8.7%; Score 234; DB 1; Length 1935;
Best Local Similarity 22.3%; Pred. No. 0.019;
Matches 126; Conservative 101; Mismatches 177; Indels 162; Gaps 25;
Qy 8 BAGPGGAQEPSPQAPAV-----BAGPGSSQAPRKEGACARTAQSGALRDVSEELS 60
Db 1363 KANSEVAQWRTKYETDAIORTLELEAKKKLQRLQDAEEVAEVAENAKSSLEKTKHRLQ 1422
Qy 61 RQLEDILSTYCVDMNQGGEDGAGQEPAPBDAKSRITYVARNGEPEPTPVVYGRKEPS 120
Db 1423 NEIEDLM-----VDVERSNAAAAAL-----DKQR 1447
Qy 121 KGPNTERTROSDEVDGRHRRPOEKKKAGLKEITLM-----QTLNTLSTPEKLAAL 176
Db 1448 NFDKILAEWKQYEEQSLESSQ--KEARSLSTFLKLNAYEESLEHLETSKRKNKL 1505
Qy 177 CKYVAILLEHRNSQOMKLLQKQSQVQEKDHLRG-----EHSKAVLARSKE- 226
Db 1506 QEBISDLTEQLGSSGKTIHELEKVKQLAEKLEQLQSALEESSEASLEHEEGKILRAQLEF 1565
Qy 227 -----SLCRLQHNRL-----KEEG----- 243
Db 1566 NQIKAEWKRLAEKDEBEQAKENHLRVVDSLQTSLSDAETSRSEALRVKKQMEGDLNEM 1625
Qy 244 -VQRA--REBEKKREVTSHPOVTINDIQLOMEQHNERNKSLROENMELAE- 292
Db 1626 EIQLSHANNWAAQAQKQVKS-LQSLKDTQIQLDVAVRNDLLK-ENIAIVERNNLQA 1683
Qy 293 ----LKLLEQYE----LREHIDKVFKNKL-----QQQLVDAKLOQAQEMKE 334
Db 1684 ELBELRAVVEQTERSRLAEQELIETSERVQLLHQSNTSLINQKKMEADLSQLOTEVEE 1743
Qy 335 A--EESHOREKDFLLKEAVESQWMC--ELMKQOET--HLKQQLALVTEKEPEEQNTLSKS 388
Db 1744 AVQESRNAEEK--AKKAITDAAMAEELKKEQUTSAHLERKKNMEQTIKIQHRLDEA 1800
Qy 389 SEV-FTTFKQEMKMTYKIKKLEKETTMYRSRVSESSNKALLEMAEKTVEDKE-LEGLOV 446
Db 1801 EQIALGKGKQKQLAEARVRELENE-----LE-AEQK--RNAESVKGMRK 1842
RESULT 31
SCPI_MESAU STANDARD; PRT; 845 AA.
AC Q60563;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Synaptonemal complex protein 1 (SCP-1 protein) (Meiotic chromosome synaptonemal complex protein) (Fragment).
DE SCPI OR SYN1.
OS Mesocricetus auratus (Golden hamster).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Cricetinae;
OC Mesocricetus.
OX NCBI_TaxID=10036;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=testis;
RX MEDLINE=95181577; PubMed=7876343;
RA Dobson M.J., Pearlman R.E., Karaiskakis A., Spyropoulos B., Moens P.B.;
RA "Synaptonemal complex proteins: occurrence, epitope mapping and chromosome disjunction."
RL J. Cell Sci 107:2749-2760(1994).
CC -!- FUNCTION: Major component of the transverse filaments of

CC synaptonemal complexes (SCS), formed between homologous
CC chromosomes during meiotic prophase (By similarity). Has non-
CC specific DNA binding capability.
CC
CC -|- SUBCELLULAR LOCATION: Nuclear. In tripartite segments of
CC synaptonemal complexes, between lateral elements in the nucleus.
CC Found only where the chromosome cores are synapsed. Its N-terminus
CC is found towards the centre of the synaptonemal complex while the
CC C-terminus extends well into the lateral domain of the
CC synaptonemal complex.
CC
CC -|- DOMAIN: Consists of an alpha-helical stretch of 700 AA residues,
CC flanked by N- and C-terminal globular domains. The C-terminal
CC domain has DNA-binding capacity (By similarity).
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC
CC EMBL: L32978; AAC42039.1; -.
CC PIR: I48176; I48176.
CC InterPro: IPR008827; SCP-1.
CC Pfam: PF05493; SCP-1; 1.
CC
CC Nuclear protein; Meiosis; Cell division; Phosphorylation;
CC DNA-binding; Coiled coil.
CC
CC NON TER 1
CC DOMAIN <1 672 COILED COIL (POTENTIAL).
CC DOMAIN 553 556 NUCLEAR LOCALIZATION SIGNAL (POTENTIAL).
CC DOMAIN 753 756 NUCLEAR LOCALIZATION SIGNAL (POTENTIAL).
CC DOMAIN 830 838 ARG/LYS-RICH (BASIC).
CC SEQUENCE 845 AA; 99401 MW; D7F28873C824C6A8 CRC64;
CC
CC Query Match 8.6%; Score 233.5; DB 1; Length 845;
CC Best Local Similarity 24.0%; Pred. No. 0.0091;
CC Matches 109; Conservative 86; Mismatches 168; Indels 91; Gaps 19;
CC
Cc 90 EPEDAKSRVTYVARGPEPTPVYGSKEPSKGPNTVEERQSDEVCGRD----- 139
Cc 54 EYEEETRVQVYVDNNNIKRMILAF-----BELRVQANRLDMHFKLKEDHE 101
Cc 140 ---HRRPQKKAGLGGKITLMTQTLNTLSTPEKLAALCKKYAELEBHRNSQKQ-- 194
Cc 102 KIQHLOEYKKEVNDKENQVSL---LIQTEKENKM---KDLTFLESDRKNVQLD 154
Cc 195 ---KLQKKQSOLVQKQHLRG--BHSKAVLARS--KLESICRELQNRSLKEGVQAR 248
Cc 155 KTKLQDENVKELNKKKHDLTSELDTRMSLQSRMNTQKALEEDLIQATKTIYQUTEKBA 214
Cc 249 BEERKKEVTSF-FOV-----TLNDIOLQMEQHN--ERNSKLRQENMELAERLKLI 297
Cc 215 QMEFPNKAQTHDSFWTELKATCTTBEL-LRTQQRVLVKNEDQLKILTWELQKSNELD 273
Cc 298 EGYELREHDKVPKHKDL---QQQLVDKIQQAQEMLKRAERHQ-----REKDF- 345
Cc 274 EMTKFKNNNEVKLEELKKILAEQDKLADK-KQVEKLAELQGEQBELTLILLOTREKEVH 332
Cc 346 -----LKAEVSRMCELMKQOETHLQKQ-----LALYTEKFEFQFQ--- 383
Cc 333 DLERQLLVTKISDQNYSKQVBELTKKEEKKLKNALETASCKSLNNKLTQTNDMAL 392
Cc 384 TLKSKSVFTTFKQEMEKQTKIKKLEKETTYSRWNSSNKALAEAE-KTVRDKBLE 442
Cc 393 ELAKYQEDITNSKQOERMLQKQENLEBKETHLRLDELESVEKFTIQGNEVKCKLDKSEE 452
Cc 443 ---GLQVKIQLEKLCRALQTERNDLNKRVODLS 473
Cc 453 NARSIECEVLKCKQMKILNKNNKLNKQAEKNS 486

RESULT 32
MYH2_HUMAN

ID MYH2 HUMAN STANDARD; PRT; 1941 AA.
AC Q9UXK2; Q14322; Q16229;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DE 15-MAR-2004 (Rel. 43, Last annotation update)
DE Myosin heavy chain, skeletal muscle, adult 2 (Myosin heavy chain Ila)
DE (MYHC-IIa).
GN MYH2 OR MYHSA2.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OC NCB TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Skeletal muscle;
RX MEDLINE=99318869; PubMed=10388558;
RA Weiss A., Schaffino S., Leinwand L.A.;
RT "Comparative sequence analysis of the complete human sarcomeric myosin
RT heavy chain family: implications for functional diversity";
RL J. Mol. Biol. 290:61-75(1999).
RN [2]
RP SEQUENCE OF 1711-1941 FROM N.A.
RC TISSUE=Skeletal muscle;
RX MEDLINE=95109625; PubMed=7545970;
RA Smerdu V., Karsch-Mizrachi I., Campione M., Leinwand L.,
RA Schaffino S.;
RT "Type IIX myosin heavy chain transcripts are expressed in type IIB
RT fibers of human skeletal muscle";
RL Am. J. Physiol. 267:C1723-C1728(1994).
RN [3]
RP SEQUENCE OF 1823-1941 FROM N.A.
RC TISSUE=Skeletal muscle;
RX MEDLINE=95270723; PubMed=7751403;
RA Emmon S., Sant'ana Pereira J., Sargeant T., Young A., Goldspink G.;
RT "Characterization of human skeletal muscle fibres according to the
RT myosin heavy chains they express";
RL J. Muscle Res. Cell Motil. 16:35-43(1995).
CC -|- FUNCTION: Muscle contraction.
CC -|- SUBUNIT: Muscle myosin is a hexameric protein that consists of 2
CC heavy chain subunits (MHC), 2 alkali light chain subunits (MLC)
CC and 2 regulatory light chain subunits of the myofibrils.
CC -|- SUBCELLULAR LOCATION: Thick filaments of the myofibrils.
CC -|- DOMAIN: The rodlike tail sequence is highly repetitive, showing
CC cycles of a 28-residue repeat pattern composed of 4 heptapeptides,
CC characteristic for alpha-helical coiled coils.
CC -|- MISCELLANEOUS: Each myosin heavy chain can be split into 1 light
CC meromyosin (LMW) and 1 heavy meromyosin (HMM). It can later be
CC split further into 2 globular subfragments (S1) and 1 rod-shaped
CC subfragment (S2).
CC -|- SIMILARITY: Contains 1 myosin-like globular head domain.
CC -|- SIMILARITY: Contains 1 IQ domain.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC
CC EMBL: AF111784; AAD29950.1; -.
CC EMBL: S73840; AAC3916.1; -.
CC EMBL: Z32858; CAAB3687.1; -.
CC PIR: I51912; I51912.
CC HSSP: P13538; 2MYS.
CC Genew: HGNC:7572; MYH2.
CC MTM; 160740; -.
CC GO: 0005859; C:muscle myosin; TAS.
CC GO: 0003779; F:actin binding; NAS.
CC GO: 0005524; F:ATP binding; NAS.
CC GO: 0005516; F:calmodulin binding; NAS.
CC GO: 0003776; F:muscle motor activity; TAS.
CC GO: 0006936; P:muscle contraction; TAS.

DR	InterPro; IPR000048; IQ region.	DR	DE	15-MAR-2004 (Rel. 43, Last annotation update)
DR	InterPro; IPR001609; Myosin head.	DR	GN	Myosin heavy chain, cardiac muscle beta isoform (MyHC-beta).
DR	InterPro; IPR004009; Myosin N.	DR	MH7.	
DR	InterPro; IPR002928; Myosin tail.	OS	Rattus norvegicus (Rat).	
DR	Pfam; PF00612; IQ; 2.	OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;	
DR	Pfam; PF00063; myosin head; 1.	OC	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.	
DR	Pfam; PF02736; Myosin_N; 1.	OX	NCBI_TaxID=10116;	
DR	Pfam; PF01576; Myosin_tail; 1.	RN	[1]	SEQUENCE FROM N.A.
DR	PRINTS; PR00193; MYOSINHEAVY.	RP	TISSUE=Heart;	
DR	PRODOM; PD000355; myosin head; 1.	RC	MEDLINE=90016823; PubMed=2798112;	
DR	SMART; SM00015; IQ; 1.	RX	Kraft R., Bravo-Zehnder M., Taylor D., Leinwand L.A.;	
DR	SMART; SM00242; MYSc; 1.	RA	"Complete nucleotide sequence of full length cDNA for rat beta	
DR	PROSITE; PS50096; IQ; 1.	RT	cardiac myosin heavy chain."	
KW	Myosin: Muscle protein; Coiled coil; Thick filament; Actin-binding;	RL	Nucleic Acids Res. 17:7529-7530(1989).	
KW	Calmodulin-binding; ATP-binding; Methylation; Multigene family.	RN	[2]	DISCUSSION OF SEQUENCE.
FT	DOMAIN 1 786	RP	MEDLINE=90133919; PubMed=2614840;	
FT	DOMAIN 787 816	RX	McNally E.M., Kraft R., Bravo-Zehnder M., Taylor D., Leinwand L.A.;	
FT	DOMAIN 845 1941	RA	"Full-length rat alpha and beta cardiac myosin heavy chain sequences.	
FT	NP_BIND 179 186	RT	Comparisons suggest a molecular basis for functional differences."	
FT	CONFLICT 1844 1844	RL	J. Mol. Biol. 210:665-671(1989).	
SQ	SEQUENCE 1941 AA; 223043 MW; 681E866F83AE83F CRC64;	RN	[3]	SEQUENCE OF 1524-1935 FROM N.A.
		RP	MEDLINE=82220036; PubMed=7045682;	
		RX	Mahdavi V., Periasamy M., Nadal-Ginard B.;	
		RA	"Molecular characterization of two myosin heavy chain genes expressed	
		RT	in the adult heart."	
		RL	Nature 297:659-664(1982).	
		RN	[4]	SEQUENCE OF 1871-1935 FROM N.A.
		RP	STRAIN=Wistar; TISSUE=Heart;	
		RX	MEDLINE=85179510; PubMed=6241892;	
		RA	Mahdavi V., Lompre A.M., Chambers A.P., Nadal-Ginard B.;	
		RT	"Cardiac myosin heavy chain isozymic transitions during development	
		RL	and under pathological conditions are regulated at the level of mRNA	
		RN	availability."	
		RP	Bur. Heart J. 5:181-191(1984).	
		RX	FUNCTION: Muscle contraction.	
		RA	SUBUNIT: Muscle myosin is a hexameric protein that consists of 2	
		RT	heavy chain subunits (MHC), 2 alkali light chain subunits (MLC)	
		RL	and 2 regulatory light chain subunits (MLC-2).	
		RN	SUBCELLULAR LOCATION: Thick filaments of the myofibrils.	
		RP	DOMAIN: The rodlike tail sequence is highly repetitive, showing	
		RX	cycles of a 28-residue repeat pattern composed of 4 heptapeptides,	
		RA	characteristic for alpha-helical coiled coils.	
		RT	MISCELLANEOUS: Each myosin heavy chain can be split into 1 light	
		RL	meromyosin (LMW) and 1 heavy meromyosin (HMW). It can later be	
		RN	split further into 2 globular subfragments (S1) and 1 rod-shaped	
		RP	subfragment (S2).	
		RX	MISCELLANEOUS: The cardiac alpha isoform is a 'fast' ATPase	
		RA	myosin, while the beta isoform is a 'slow' ATPase.	
		RT	SIMILARITY: Contains 1 myosin-like globular head domain.	
		RL	SIMILARITY: Contains 1 IQ domain.	
		RN	THIS SWISS-PROT entry is copyright. It is produced through a collaboration	
		RP	between the Swiss Institute of Bioinformatics and the EMBL outstation -	
		RX	the European Bioinformatics Institute. There are no restrictions on its	
		RA	use by non-profit institutions as long as its content is in no way	
		RT	modified and this statement is not removed. Usage by and for commercial	
		RL	entities requires a license agreement (See http://www.isb-sib.ch/announce/	
		RN	or send an email to license@isb-sib.ch).	
		RP	EMBL; X15939; CAA34065.1; -	
		RX	EMBL; J00752; AAA41654.1; -	
		RA	EMBL; M32698; AAA41659.1; -	
		RT	PIR; S06006; S06006.	
		RL	HSP; P08799; 1MND.	
		RN	InterPro; IPR000048; IQ_region.	
		RP	InterPro; IPR001609; myosin head.	
		RX	InterPro; IPR004009; Myosin_N.	
		RA	InterPro; IPR002928; Myosin_tail.	
		RT	Pfam; PF00612; IQ; 2.	

DR	InterPro; IPR000048; IQ region.	DR	DE	15-MAR-2004 (Rel. 43, Last annotation update)
DR	InterPro; IPR001609; Myosin head.	DR	GN	Myosin heavy chain, cardiac muscle beta isoform (MyHC-beta).
DR	InterPro; IPR004009; Myosin N.	DR	MH7.	
DR	InterPro; IPR002928; Myosin tail.	OS	Rattus norvegicus (Rat).	
DR	Pfam; PF00612; IQ; 2.	OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;	
DR	Pfam; PF00063; myosin head; 1.	OC	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.	
DR	Pfam; PF02736; Myosin_N; 1.	OX	NCBI_TaxID=10116;	
DR	Pfam; PF01576; Myosin_tail; 1.	RN	[1]	SEQUENCE FROM N.A.
DR	PRINTS; PR00193; MYOSINHEAVY.	RP	TISSUE=Heart;	
DR	PRODOM; PD000355; myosin head; 1.	RC	MEDLINE=90016823; PubMed=2798112;	
DR	SMART; SM00015; IQ; 1.	RX	Kraft R., Bravo-Zehnder M., Taylor D., Leinwand L.A.;	
DR	SMART; SM00242; MYSc; 1.	RA	"Complete nucleotide sequence of full length cDNA for rat beta	
DR	PROSITE; PS50096; IQ; 1.	RT	cardiac myosin heavy chain."	
KW	Myosin: Muscle protein; Coiled coil; Thick filament; Actin-binding;	RL	Nucleic Acids Res. 17:7529-7530(1989).	
KW	Calmodulin-binding; ATP-binding; Methylation; Multigene family.	RN	[2]	DISCUSSION OF SEQUENCE.
FT	DOMAIN 1 786	RP	MEDLINE=90133919; PubMed=2614840;	
FT	DOMAIN 787 816	RX	McNally E.M., Kraft R., Bravo-Zehnder M., Taylor D., Leinwand L.A.;	
FT	DOMAIN 845 1941	RA	"Full-length rat alpha and beta cardiac myosin heavy chain sequences.	
FT	NP_BIND 179 186	RT	Comparisons suggest a molecular basis for functional differences."	
FT	CONFLICT 1844 1844	RL	J. Mol. Biol. 210:665-671(1989).	
SQ	SEQUENCE 1941 AA; 223043 MW; 681E866F83AE83F CRC64;	RN	[3]	SEQUENCE OF 1524-1935 FROM N.A.
		RP	MEDLINE=82220036; PubMed=7045682;	
		RX	Mahdavi V., Periasamy M., Nadal-Ginard B.;	
		RA	"Molecular characterization of two myosin heavy chain genes expressed	
		RT	in the adult heart."	
		RL	Nature 297:659-664(1982).	
		RN	[4]	SEQUENCE OF 1871-1935 FROM N.A.
		RP	STRAIN=Wistar; TISSUE=Heart;	
		RX	MEDLINE=85179510; PubMed=6241892;	
		RA	Mahdavi V., Lompre A.M., Chambers A.P., Nadal-Ginard B.;	
		RT	"Cardiac myosin heavy chain isozymic transitions during development	
		RL	and under pathological conditions are regulated at the level of mRNA	
		RN	availability."	
		RP	Bur. Heart J. 5:181-191(1984).	
		RX	FUNCTION: Muscle contraction.	
		RA	SUBUNIT: Muscle myosin is a hexameric protein that consists of 2	
		RT	heavy chain subunits (MHC), 2 alkali light chain subunits (MLC)	
		RL	and 2 regulatory light chain subunits (MLC-2).	
		RN	SUBCELLULAR LOCATION: Thick filaments of the myofibrils.	
		RP	DOMAIN: The rodlike tail sequence is highly repetitive, showing	
		RX	cycles of a 28-residue repeat pattern composed of 4 heptapeptides,	
		RA	characteristic for alpha-helical coiled coils.	
		RT	MISCELLANEOUS: Each myosin heavy chain can be split into 1 light	
		RL	meromyosin (LMW) and 1 heavy meromyosin (HMW). It can later be	
		RN	split further into 2 globular subfragments (S1) and 1 rod-shaped	
		RP	subfragment (S2).	
		RX	MISCELLANEOUS: The cardiac alpha isoform is a 'fast' ATPase	
		RA	myosin, while the beta isoform is a 'slow' ATPase.	
		RT	SIMILARITY: Contains 1 myosin-like globular head domain.	
		RL	SIMILARITY: Contains 1 IQ domain.	
		RN	THIS SWISS-PROT entry is copyright. It is produced through a collaboration	
		RP	between the Swiss Institute of Bioinformatics and the EMBL outstation -	
		RX	the European Bioinformatics Institute. There are no restrictions on its	
		RA	use by non-profit institutions as long as its content is in no way	
		RT	modified and this statement is not removed. Usage by and for commercial	
		RL	entities requires a license agreement (See http://www.isb-sib.ch/announce/	
		RN	or send an email to license@isb-sib.ch).	
		RP	EMBL; X15939; CAA34065.1; -	
		RX	EMBL; J00752; AAA41654.1; -	
		RA	EMBL; M32698; AAA41659.1; -	
		RT	PIR; S06006; S06006.	
		RL	HSP; P08799; 1MND.	
		RN	InterPro; IPR000048; IQ_region.	
		RP	InterPro; IPR001609; myosin head.	
		RX	InterPro; IPR004009; Myosin_N.	
		RA	InterPro; IPR002928; Myosin_tail.	
		RT	Pfam; PF00612; IQ; 2.	

DR	InterPro; IPR000048; IQ region.	DR	DE	15-MAR-2004 (Rel. 43, Last annotation update)
DR	InterPro; IPR001609; Myosin head.	DR	GN	Myosin heavy chain, cardiac muscle beta isoform (MyHC-beta).
DR	InterPro; IPR004009; Myosin N.	DR	MH7.	
DR	InterPro; IPR002928; Myosin tail.	OS	Rattus norvegicus (Rat).	
DR	Pfam; PF00612; IQ; 2.	OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;	
DR	Pfam; PF00063; myosin head; 1.	OC	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.	
DR	Pfam; PF02736; Myosin_N; 1.	OX	NCBI_TaxID=10116;	
DR	Pfam; PF01576; Myosin_tail; 1.	RN	[1]	SEQUENCE FROM N.A.
DR	PRINTS; PR00193; MYOSINHEAVY.	RP	TISSUE=Heart;	
DR	PRODOM; PD000355; myosin head; 1.	RC	MEDLINE=90016823; PubMed=2798112;	
DR	SMART; SM00015; IQ; 1.	RX	Kraft R., Bravo-Zehnder M., Taylor D., Leinwand L.A.;	
DR	SMART; SM00242; MYSc; 1.	RA	"Complete nucleotide sequence of full length cDNA for rat beta	
DR	PROSITE; PS50096; IQ; 1.	RT	cardiac myosin heavy chain."	
KW	Myosin: Muscle protein; Coiled coil; Thick filament; Actin-binding;	RL	Nucleic Acids Res. 17:7529-7530(1989).	
KW	Calmodulin-binding; ATP-binding; Methylation; Multigene family.	RN	[2]	DISCUSSION OF SEQUENCE.
FT	DOMAIN 1 786	RP	MEDLINE=90133919; PubMed=2614840;	
FT	DOMAIN 787 816	RX	McNally E.M., Kraft R., Bravo-Zehnder M., Taylor D., Leinwand L.A.;	
FT	DOMAIN 845 1941	RA	"Full-length rat alpha and beta cardiac myosin heavy chain sequences.	
FT	NP_BIND 179 186	RT	Comparisons suggest a molecular basis for functional differences."	
FT	CONFLICT 1844 1844	RL	J. Mol. Biol. 210:665-671(1989).	
SQ	SEQUENCE 1941 AA; 223043 MW; 681E866F83AE83F CRC64;	RN	[3]	SEQUENCE OF 1524-1935 FROM N.A.
		RP	MEDLINE=82220036; PubMed=7045682;	
		RX	Mahdavi V., Periasamy M., Nadal-Ginard B.;	
		RA	"Molecular characterization of two myosin heavy chain genes expressed	
		RT	in the adult heart."	
		RL	Nature 297:659-664(1982).	
		RN	[4]	SEQUENCE OF 1871-1935 FROM N.A.
		RP	STRAIN=Wistar; TISSUE=Heart;	
		RX	MEDLINE=85179510; PubMed=6241892;	
		RA	Mahdavi V., Lompre A.M., Chambers A.P., Nadal-Ginard B.;	
		RT	"Cardiac myosin heavy chain isozymic transitions during development	
		RL	and under pathological conditions are regulated at the level of mRNA	
		RN	availability."	
		RP	Bur. Heart J. 5:181-191(1984).	
		RX	FUNCTION: Muscle contraction.	
		RA	SUBUNIT: Muscle myosin is a hexameric protein that consists of 2	
		RT	heavy chain subunits (MHC), 2 alkali light chain subunits (MLC)	
		RL	and 2 regulatory light chain subunits (MLC-2).	
		RN	SUBCELLULAR LOCATION: Thick filaments of the myofibrils.	
		RP	DOMAIN: The rodlike tail sequence is highly repetitive, showing	
		RX	cycles of a 28-residue repeat pattern composed of 4 heptapeptides,	
		RA	characteristic for alpha-helical coiled coils.	
		RT	MISCELLANEOUS: Each myosin heavy chain can be split into 1 light	
		RL	meromyosin (LMW) and 1 heavy meromyosin (HMW). It can later be	
		RN	split further into 2 globular subfragments (S1) and 1 rod-shaped	
		RP	subfragment (S2).	
		RX	MISCELLANEOUS: The cardiac alpha isoform is a 'fast' ATPase	
		RA	myosin, while the beta isoform is a 'slow' ATPase.	
		RT	SIMILARITY: Contains 1 myosin-like globular head domain.	
		RL	SIMILARITY: Contains 1 IQ domain.	
		RN	THIS SWISS-PROT entry is copyright. It is produced through a collaboration	
		RP	between the Swiss Institute of Bioinformatics and the EMBL outstation -	
		RX	the European Bioinformatics Institute. There are no restrictions on its	
		RA	use by non-profit institutions as long as its content is in no way	
		RT	modified and this statement is not removed. Usage by and for commercial	
		RL	entities requires a license agreement (See http://www.isb-sib.ch/announce/	
		RN	or send an email to license@isb-sib.ch).	
		RP	EMBL; X15939; CAA34065.1; -	
		RX	EMBL; J00752; AAA41654.1; -	
		RA	EMBL; M32698; AAA41659.1; -	
		RT	PIR; S06006; S06006.	
		RL	HSP; P08799; 1MND.	
		RN	InterPro; IPR000048; IQ_region.	
		RP	InterPro; IPR001609; myosin head.	
		RX	InterPro; IPR004009; Myosin_N.	
		RA	InterPro; IPR002928; Myosin_tail.	
		RT	Pfam; PF00612; IQ; 2.	

DR	InterPro; IPR000048; IQ region.	DR	DE	15-MAR-2004 (Rel. 43, Last annotation update)
DR	InterPro; IPR001609; Myosin head.	DR	GN	Myosin heavy chain, cardiac muscle beta isoform (MyHC-beta).
DR	InterPro; IPR004009; Myosin N.	DR	MH7.	
DR	InterPro; IPR002928; Myosin tail.	OS	Rattus norvegicus (Rat).	
DR	Pfam; PF00612; IQ; 2.	OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;	
DR	Pfam; PF00063; myosin head; 1.	OC	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.	
DR	Pfam; PF02736; Myosin_N; 1.	OX	NCBI_TaxID=10116;	
DR	Pfam; PF01576; Myosin_tail; 1.	RN	[1]	SEQUENCE FROM N.A.
DR	PRINTS; PR00193; MYOSINHEAVY.	RP	TISSUE=Heart;	
DR	PRODOM; PD000355; myosin head; 1.	RC	MEDLINE=90016823; PubMed=2798112;	
DR	SMART; SM00015; IQ; 1.	RX	Kraft R., Bravo-Zehnder M., Taylor D., Leinwand L.A.;	
DR	SMART; SM00242; MYSc; 1.	RA	"Complete nucleotide sequence of full length cDNA for rat beta	
DR	PROSITE; PS50096; IQ; 1.	RT	cardiac myosin heavy chain."	
KW	Myosin: Muscle protein; Coiled coil; Thick filament; Actin-binding;	RL	Nucleic Acids Res. 17:7529-7530(1989).	
KW	Calmodulin-binding; ATP-binding; Methylation; Multigene family.	RN	[2]	DISCUSSION OF SEQUENCE.
FT	DOMAIN 1 786	RP	MEDLINE=90133919; PubMed=2614840;	
FT	DOMAIN 787 816	RX	McNally E.M., Kraft R., Bravo-Zehnder M., Taylor D., Leinwand L.A.;	
FT	DOMAIN 845 1941	RA	"Full-length rat alpha and beta cardiac myosin heavy chain sequences.	
FT	NP_BIND 179 186	RT	Comparisons suggest a molecular basis for functional differences."	
FT	CONFLICT 1844 1844	RL	J. Mol. Biol. 210:665-671(1989).	
SQ	SEQUENCE 1941 AA; 223043 MW; 681E866F83AE83F CRC64;	RN	[3]	SEQUENCE OF 1524-1935 FROM N.A.
		RP	MEDLINE=82220036; PubMed=7045682;	
		RX	Mahdavi V., Periasamy M., Nadal-Ginard B.;	
		RA	"Molecular characterization of two myosin heavy chain genes expressed	
		RT	in the adult heart."	
		RL	Nature 297:659-664(1982).	
		RN	[4]	SEQUENCE OF 1871-1935 FROM N.A.
		RP	STRAIN=Wistar; TISSUE=Heart;	
		RX	MEDLINE=85179510; PubMed=6241892;	
		RA	Mahdavi V., Lompre A.M., Chambers A.P., Nadal-Ginard B.;	
		RT	"Cardiac myosin heavy chain isozymic transitions during development	
		RL	and under pathological conditions are regulated at the level of mRNA	
		RN	availability."	
		RP	Bur. Heart J. 5:181-191(1984).	
		RX	FUNCTION: Muscle contraction.	
		RA	SUBUNIT: Muscle myosin is a hexameric protein that consists of 2	
		RT	heavy chain subunits (MHC), 2 alkali light chain subunits (MLC)	
		RL	and 2 regulatory light chain subunits (MLC-2).	
		RN	SUBCELLULAR LOCATION: Thick filaments of the myofibrils.	
		RP	DOMAIN: The rodlike tail sequence is highly repetitive, showing	
		RX	cycles of a 28-residue repeat pattern composed of 4 heptapeptides,	
		RA	characteristic for alpha-helical coiled coils.	
		RT	MISCELLANEOUS: Each myosin heavy chain can be split into 1 light	
		RL	meromyosin (LMW) and 1 heavy meromyosin (HMW). It can later be	
		RN	split further into 2 globular subfragments (S1) and 1 rod-shaped	
		RP	subfragment (S2).	
		RX	MISCELLANEOUS: The cardiac alpha isoform is a 'fast' ATPase	
		RA	myosin, while the beta isoform is a 'slow' ATPase.	
		RT	SIMILARITY: Contains 1 myosin-like globular head domain.	
		RL	SIMILARITY: Contains 1 IQ domain.	
		RN	THIS SWISS-PROT entry is copyright. It is produced through a collaboration	

DR Pfam; PF00063; myosin head; 1.
DR Pfam; PF02736; Myosin_N; 1.
DR Pfam; PF01576; Myosin_tail; 1.
DR PRINTS; PR00193; MYOSINHEAVY.
DR ProDom; PD000355; myosin_head; 1.
DR SMART; SM00015; IQ; 1.
DR SMART; SM00242; MYSC; 1.
DR PROSITE; PS50096; IQ; 1.
KW Myosin; Muscle protein; Coiled coil; Thick filament; Actin-binding;
KW ATP-binding; Methylation; Multigene family; Calmodulin-binding.
MYOSIN HEAD-LIKE.
FT DOMAIN 1 780
FT DOMAIN 781 810 IQ
FT DOMAIN 840 1935
FT NP_BIND 178 185
FT DOMAIN 655 677
FT DOMAIN 757 771
FT MOD_RES 129 129
FT CONFLICT 1529 1531
FT CONFLICT 1731 1731
FT CONFLICT 1784 1784
FT CONFLICT 1851 1851
FT CONFLICT 1858 1858
SQ SEQUENCE 1935 AA; 223082 MW; C8376C324A7BD82B CRC64;

Query Match 8.6%; Score 232; DB 1; Length 1935;
Best Local Similarity 22.3%; Pred. No. 0.023;
Matches 116; Conservative 81; Mismatches 189; Indels 134; Gaps 16;

QY 25 VEAEGPGSQAPKPEGAQRTAQSALRDVSELSQLEDLISTVCVNNQGGPGDGA 84
DB 1094 IEDEQALGSLQKXLEQARI-----EELSELE----- 1123

QY 85 QGPAPPEDEKERTVVARNGPEPTVVVGEKPSKGDPTBEIRQSDVGVDRHRRPQ 144
DB 1124 -AERTARAKVEKRSLSRELEB-----ISELEAGGATSVQI-----EMKKREA 1169

QY 145 EKKAKGLGKEITLIMQTLNTLSTPEKLAALCKY-----AELESHRSQKMKLLQKK 200
DB 1170 EFQKRRDLSEATL-----QHEATAAALKKHADSVAELGEQIDMLQRYKQKLEKE 1220

QY 201 QSQVQEKQHLGHEHSAVLARSKLSLCT-----EL 232
DB 1221 KSPFKLEDDVTSNMEQIIRKAKANLEKCKTLEDQWNEHRSKABETORSVNDLTORAKL 1280

QY 233 QRHN-----RSLKE-----EGVQARBEERKKEVTSF-FQVTLNDI 268
DB 1281 QTENGELSRQDEKEALISQITRGKLTYYTQLEDLKRQLEEEVKAKNALAHQSAHDHC 1340

QY 269 QLOMEQHNEHNSK-----LRQNNELAEKIKLIEQYELREHIDKVKFKHQLQQLV 321
DB 1341 DLLREQYEETBAKELQRYLSKANSEVAQWRKTYETDAIORTTELEEA--KKGLAQRLQ 1398

QY 322 DAKLQQAQE-----MLKEAERHOREKDFLLKEAVESQRMCELMKQQTTHLKQOLALY 374
DB 1399 DA--EBAVEAVNAKCSSEKTEKHLQNEIEDLAVDVERSNAAAALDKKQRFDKILVEN 1456

QY 375 TEKPEPQNTLSKSEVFTTFKQEMERQTKKIKLEKFTTNYRSRVSSNKALLEMAEEK 434
DB 1457 KQYEESSQSESSQKEARSLSLTEL-----FLKNAYEESLEHLETFKREKNLQEBISDL 1512

QY 435 TVRDKEGLEGLQVKIQRLEKLCRALOTERNLNNKRVODLSA 474
DB 1513 T---EQLGSTGKSIHELEKTKRQLEAEKLELSQNALAEAA 1549

RESULT 34
RA50_PVRFU
ID -RA50_PVRFU STANDARD; PRT; 882 AA.
AC PE9301;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE DNA double-strand break repair rad50 ATPase.

GN RAD50 OR PFI167.
OS Pyrococcus furiosus.
OC Archaea; Euryarchaeota; Thermococci; Thermococcales; Thermococcaceae;
OC Pyrococcus.
OX NCBI_TaxID=2261;
EN [1]
RP SEQUENCE FROM N.A., AND CHARACTERIZATION.
RC STRAIN=Vc1 / DSM 3638 / ATCC 43587 / JCM 8422;
RX MEDLINE=20485553; PubMed=11029422;
RA Hopfner K.-P., Karcher A., Shin D., Fairley C., Tainer J.A.,
Carney J.P.,
RT "Mrell and Rad50 from Pyrococcus furiosus: cloning and biochemical
characterization reveal an evolutionarily conserved multi-protein
machine";
RL J. Bacteriol. 182:6036-6041(2000).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=Vc1 / DSM 3638 / ATCC 43587 / JCM 8422;
RX Weiss R.B., Dunn D.M., Robb F.T., Brown J.R.;
RT "The complete sequence of the Pyrococcus furiosus genome.";
RL Submitted (FEB-2002) to the EMBL/GenBank/DBJ databases.
RN [3]
RP X-RAY CRYSTALLOGRAPHY (1.6 ANGSTROMS) OF 1-149.
RC STRAIN=Vc1 / DSM 3638 / ATCC 43587 / JCM 8422;
RX Hopfner K.-P., Karcher A., Shin D.S., Craig L., Arthur L.M.,
Carney J.P., Tainer J.A.;
RT "Structural biology of Rad50 ATPase: ATP-driven conformational
control in DNA double-strand break repair and the ABC-ATPase
superfamily";
RL Cell 101:789-800(2000).
CC -!- FUNCTION: Involved in DNA double-strand break repair (DSBR). The
rad50/mrell complex possesses single-strand endonuclease activity
and ATP-dependent double-strand-specific exonuclease activity.
CC Rad50 provides an ATP-dependent control of mrell by unwinding
and/or repositioning DNA ends into the mrell active site.
CC -!- SUBUNIT: Forms a complex with mrell.
CC -!- SIMILARITY: Belongs to the SMC family. RAD50 subfamily.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
or send an email to license@isb-sib.ch).
CC
CC EMBL; AB010225; AAL81291.1; -.
DR PDB; 1P2T; 20-SEP-00.
DR PDB; 1P2U; 02-AUG-00.
DR PDB; 1I18; 30-MAY-01.
DR PDB; 1L8D; 28-AUG-02.
DR HAMAP; MF_00449; -; 1.
DR InterPro; IPR003439; ABC transporter.
DR InterPro; IPR007517; Rad50 zn_hook.
DR Pfam; PF04423; Rad50 zn_hook; 1.
DR Pfam; PF02463; SMC_N; 1.
DR ProDom; PD000006; ABC transporter; 1.
KW DNA repair; Hydrolyase; ATP-binding; Coiled coil; 3D-structure;
KW Complete proteome.
FT NP_BIND 30 37 ATP.
FT DOMAIN 148 744
FT SEQUENCE 882 AA; 103839 MW; 3ADCB250382A99E CRC64;

Query Match 8.6%; Score 231.5; DB 1; Length 882;
Best Local Similarity 23.3%; Pred. No. 0.011;
Matches 103; Conservative 90; Mismatches 148; Indels 101; Gaps 14;

QY 127 EBIROSDEVGDRHRRPOEKKAKGLGKEITLIMQTLNTLSTPEKLAALCKYAELEE 186
DB 144 DAILESDEAREKVRVLEVLNLDKFEYKLSLKXTINNRIKEYRDLARTENIEELKE 203

D5 1036 EGSLEQEKVKYRMDLE-----BAKRLGDLKLQTSB--IMDLEND-KLQLEELKKKKE 1085
QY 72 VDNNGGPGDGAQGBEPEDA-----EKSTYVAR-----NGEPEFTPVVYGEKEP 119
D5 1086 FDISQ-----QNSKIDBQALALQLOKXENQARIBELFELEAERTABAKVEKLR 1137
QY 120 SKGDPNTEIRQS-DEVG-----DRHRRPPOKKKAKGLGHEIILMLQTLNLTSTPEB 171
D5 1138 SOLTRELEETISERLEAGGATSVOIENKKBKBAEFQGRDRDLEATL-----QHEA 1188
QY 172 KLAALCKKY-----ABELEHERNSQKMKLQKQKQSOVLQKDKHLRGHSHKAVLRSKLS 227
D5 1189 TAAALRXKHADSVABELGQIDNLRVKQKLEKSEFKLELDVDTSHMEQIIKAKANLEK 1248
QY 228 LCR-----ELQRHNRSLKBEGVORAR----- 248
D5 1249 VSRTELEDOANRYKLFQEAQSRSLNFTTQAKLQTEGELARQLSEKALINQWLRKLS 1308
QY 249 -----BEEKKEKVTSH-FQVTLNDLQLOKQEHNER-----NSK 281
D5 1309 YTOOMEDLKQLEBEGKAKNALAHALQASRDCCDLLEBQTEHEENAKAEHLQVLSKANSE 1368
QY 282 LR-----QENMELAEURLKLEJOEYELREEHIDKY-----FKHK---DLQ 318
D5 1369 VAQWRKTYETDAIQRTTEELEBAKKAQRLQDAEBAEVAVNAKCSSLEKTKHRLQNEIED 1428
QY 319 QLVDAKLOQAOEMLEKABERHOREKDFLLKEAVESORMCELMKQOETHLQOOLALYTEKF 378
D5 1429 LMDVRSNA-----AAAALDKKQKQNFDKILAE--WKQKYESQSELESSQKBAESLSTELF 1483
QY 379 BEFQNTLSKSEVFTTFKQEMKNTKKI-----KKLEKETTMYR 417
D5 1484 -KLNVAEESLEHLETTFKRENKNIQBEISDLTEQLGEGGNVHELEKIRQLEVEKLELQ 1542
QY 418 SRWESSKALLEMAEKTVDKLEGLQVKIQRLKLC-----RALQTERNDLNRQVOD 471
D5 1543 SALFEA-EASLEHEEGKILR-AQLEFNQIKABIERKLAEDMEQAKRNHL--RVVD 1596

RESULT 37

RRBI_MOUSE
ID -RRBI_MOUSE STANDARD; PRT; 1605 AA.
AC Q99PL5; Q99PK5; Q99PK6; Q99PK7; Q99PK8; Q99PK9; Q99PL0; Q99PL1;
AC Q99PL2; Q99PL3; Q99PL4; Q9CS20;
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Ribosome-binding protein 1 (Ribosome receptor protein) (mRBP).
GN RRPB1.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
CX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A. (ISOFORMS RRP61/P180; RRP47; RRP41; RRP16.8;
RP RRP15A; RRP15B; RRP10; RRP5.4; RRP2; RRP1.8 AND RRP0), AND TISSUE
RP SPECIFICITY.
RC STRAIN=ICR; TISSUE=Embryo;
RX MEDLINE=21100429; PubMed=11167022;
RA Kim Y.-J., Lee M.-C., Kim S.-J., Chun J.-Y.;
RT Identification and characterization of multiple isoforms of a mouse
RT ribosome receptor.";
RL Gene 261:337-344 (2000).
RN [2]
RP SEQUENCE FROM N.A. (ISOFORM 2).
RC TISSUE=Kidney;
RX MEDLINE=22388257; PubMed=12477932;
RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA Klausner R.D., Collins P.S., Wagner L., Shenmen C.M., Schuler G.D.,
RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.P., Bhat N.K.,
RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA Stאתלton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,

Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,
RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullany S.J.,
RA Bosak S.A., McKernan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay E.J., Hulyk S.W.,
RA Villalón D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA Fahy J., Helton G., Kettner M., Madan A., Rodriguez S., Sanchez A.,
RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA Rodriguez A.C., Grimwood J., Schmutz J.J., Myers R.M., Smalhus D.E.,
RA Butterfield J.S.N., Krzywinski M.I., Skalska U., Smalhus D.E.,
RA Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.;
RT "Generation and initial analysis of more than 15,000 full-length
RT human and mouse cDNA sequences.";
RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903 (2002).
RN [3]
RP SEQUENCE OF 736-1605 FROM N.A. (ISOFORM 3).
RC STRAIN=C57BL/6J; TISSUE=Embryo;
RX MEDLINE=21085660; PubMed=11217851;
RA Kawai J., Shinagawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,
RA Arakawa T., Hara A., Fukunishi Y., Konno H., Kondo S., Yamataka I.,
RA Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamataka I.,
RA Saito T., Okazaki Y., Gojobori T., Bono H., Kasukawa T., Saito R.,
RA Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,
RA Fleischmann W., Gaasterland T., Gissi C., King B., Kochiwa H.,
RA Kuehl P., Lewis S., Matsuo Y., Nikaido I., Pesole G., Quackenbush J.,
RA Schriml L.M., Staubli F., Suzuki R., Tomita M., Wagner L., Washio T.,
RA Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barsh G.,
RA Blake J., Boffelli D., Bojunga N., Carninci P., de Bonaldo M.F.,
RA Brownstein M.J., Bulc D., Fletcher C., Fujita M., Gariboldi M.,
RA Gustincich S., Hill D., Hofmann M., Hume D.A., Kamiya M., Lee N.H.,
RA Lyons P., Marchionni L., Mashima J., Mazzarelli J., Mombauts P.,
RA Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,
RA Sasaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-F.,
RA Suzuki H., Toyooka K., Wang K.H., Weitz C., Whittaker C., Wilming L.,
RA Wynshaw-Boris A., Yoshida K., Hasegawa Y., Kawaji H., Kohetsuki S.,
RA Hayaishizaki Y.;
RT "Functional annotation of a full-length mouse cDNA collection.";
RL Nature 409:685-690 (2001).
CC -!- FUNCTION: Acts as a ribosome receptor and mediates interaction
CC between the ribosome and the endoplasmic reticulum membrane (By
CC similarity).
CC -!- SUBCELLULAR LOCATION: Type III membrane protein. Endoplasmic
CC reticulum (By similarity).
CC -!- ALTERNATIVE PRODUCTS:
CC Event=Alternative splicing; Named isoforms=13;
CC Comment=Additional isoforms seem to exist. Experimental
CC confirmation may be lacking for some isoforms;
CC Name=3;
CC IsoId=Q99PL5-1; Sequence=Displayed;
CC Name=RRP61; Synonyms=pl180;
CC IsoId=Q99PL5-2; Sequence=VSP_003963, VSP_003964;
CC Name=RRP47;
CC IsoId=Q99PL5-3; Sequence=VSP_003959, VSP_003963, VSP_003964;
CC Name=RRP41;
CC IsoId=Q99PL5-4; Sequence=VSP_003961, VSP_003963, VSP_003964;
CC Name=RRP16.8;
CC IsoId=Q99PL5-5; Sequence=VSP_003954, VSP_003962, VSP_003963,
CC VSP_003964;
CC Name=RRP15a;
CC IsoId=Q99PL5-6; Sequence=VSP_003958, VSP_003963, VSP_003964;
CC Name=RRP15b;
CC IsoId=Q99PL5-7; Sequence=VSP_003955, VSP_003963, VSP_003964;
CC Name=RRP10;
CC IsoId=Q99PL5-8; Sequence=VSP_003956, VSP_003963, VSP_003964;
CC Name=RRP5.4;
CC IsoId=Q99PL5-9; Sequence=VSP_003957, VSP_003963, VSP_003964;
CC Name=RRP2;
CC IsoId=Q99PL5-10; Sequence=VSP_003953, VSP_003963, VSP_003964;
CC Name=RRP1.8;
CC IsoId=Q99PL5-11; Sequence=VSP_003951, VSP_003963, VSP_003964;
CC Name=RRP0; Synonyms=RS130;
CC IsoId=Q99PL5-12; Sequence=VSP_003952, VSP_003963, VSP_003964;
CC Name=2;

CC IsoId=Q99PL5-13; Sequence=VSP_003960, VSP_003963, VSP_003964;
 CC -I- TISSUE SPECIFICITY: Widely expressed.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC ENBL; AF273683; AAKI1963.1; -
 CC ENBL; AF273684; AAKI1964.1; -
 CC ENBL; AF273685; AAKI1965.1; -
 CC ENBL; AF273686; AAKI1966.1; -
 CC ENBL; AF273687; AAKI1967.1; -
 CC ENBL; AF273688; AAKI1968.1; -
 CC ENBL; AF273689; AAKI1969.1; -
 CC ENBL; AF273690; AAKI1970.1; -
 CC ENBL; AF273691; AAKI1971.1; -
 CC ENBL; AF273692; AAKI1972.1; -
 CC ENBL; AF273693; AAKI1973.1; -
 CC ENBL; BC031452; AAKI1974.1; -
 CC ENBL; AK019664; BAE31939.1; -
 CC MGD; MGI:1932395; Rbppl.
 CC GO; GO:0030376; C: integral to endoplasmic reticulum membrane; ISS.
 CC GO; GO:0030376; F: ribosome receptor activity; ISS.
 CC GO; GO:0007165; P: signal transduction; ISS.
 CC InterPro: IPR007794; Rib. recp KP reg.
 CC Pfam; PF05104; Rib. recp KP reg. 1
 CC Transport; Protein transport; Translocation; Endoplasmic reticulum;
 CC Transmembrane; Repeat; Alternative splicing.
 CC DOMAIN 1 LUMENAL (POTENTIAL).
 CC TRANSMEM 8 28
 CC DOMAIN 29 1605
 CC DOMAIN 196 807
 CC FT VARSPLIC 177 789 Missing (in isoform RRP1.8).
 CC FT VARSPLIC 177 809 Missing (in isoform RRP0).
 CC FT VARSPLIC 196 787 Missing (in isoform RRP2).
 CC FT VARSPLIC 205 626 Missing (in isoform RRP1.6.8).
 CC FT VARSPLIC 205 666 Missing (in isoform RRP1.5b).
 CC FT VARSPLIC 209 720 Missing (in isoform RRP10).
 CC FT VARSPLIC 209 766 Missing (in isoform RRP5.4).
 CC FT VARSPLIC 269 730 Missing (in isoform RRP15a).
 CC FT VARSPLIC 309 449 Missing (in isoform RRP47).
 CC FT VARSPLIC 461 601 Missing (in isoform 2).
 CC FT VARSPLIC 466 666 Missing (in isoform RRP41).
 CC FT VARSPLIC 717 738 Missing (in isoform RRP16.8).
 CC FT VARSPLIC 841 842 GP -> VC (in isoform RRP61, isoform
 CC RRP1.8, isoform RRP0, isoform RRP2,
 CC isoform RRP16.8, isoform RRP15b, isoform
 CC RRP10, isoform RRP5.4, isoform RRP15a,
 CC isoform RRP47, isoform RRP41, isoform
 CC RRP16.8 and isoform 2).
 CC FT VARSPLIC 843 1605 Missing (in isoform RRP61, isoform
 CC RRP1.8, isoform RRP0, isoform RRP2,
 CC isoform RRP16.8, isoform RRP15b, isoform
 CC RRP10, isoform RRP5.4, isoform RRP15a,
 CC isoform RRP47, isoform RRP41, isoform

FT RRP16.8 and isoform 2).
 FT /FTId=VSP_003964.
 FT A -> G (IN REF. 1; AAKI1965).
 FT G -> S (IN REF. 1; AAKI1967).
 SQ SEQUENCE 1605 AA; 172880 MW; 05E0808BF10C665 CRC64;
 Query Match 8.5%; Score 231; DB 1; Length 1605;
 Best Local Similarity 20.1%; Pred. No. 0.021;
 Matches 142; Conservative 108; Mismatches 230; Indels 228; Gaps 27;
 QY 1 KSPSPQPEAGPGP-----AQERPSQAPAVEAEGPGSSQAPRKP----- 39
 DB 801 KXSEMAPAQGQKAMVQSQAPQAPQAPAKGSGKSGKGGPGPPDCDGPFLPYKTLVSTV 360
 QY 40 -----EGAAQAR-----TAQSCALRDVSEE-----LSRQL-----EDILSTYCVDMN 75
 DB 861 GSMVFSEGEAQLIELSEKTVIQTWHKATQKQDPVAILKQQLQKEKLLAT----- 914
 QY 76 QGGPGEDGAGQEPAPEDAKSRITVAR-NGE--PEPTPVYVGEKEPSK-----GDPNTESE 128
 DB 915 -----EQEDAAVAKSKRLRELNKEMASEKAKAAAGEAKVKQQLVARQETIAA 960
 QY 129 IRSQDEVGDRDHRPPOEKKKAK-----GLGKETITLMTLN----- 164
 DB 961 VQARQASVYRDVHVKVQQLQKIRTLQEQLENGPNTQLARQENSILRDALNATQSOVE 1020
 QY 165 -----TLSPTEKLAALCKKYABELEHNSOKMKLLQKQSQVQLVQSKDHLGRHESKAV 219
 DB 1021 SKQNTLAKLRQLSKVNLVEKSEASQEQQRKALEAKAATFEKVLQQLQASHKESE 1080
 QY 220 LA-RSKLESLSREL---ORHNESLAKEGVQVAREEEKKEVTSFTQVTLNDIQLQMEQH 275
 DB 1081 EALQRLREVTRELCAQTSHANLRAD-AKRAEQEQQVVAELHSLKLOSSEVSKSCEEL 1139
 QY 276 NERNESKL---ROENMELAEKLLIQEYELREBEHDKVFKHDKLQOQLVDAK----- 324
 DB 1140 SSLHGQLKEARNSQLTIRISIALLEAGQAQ-DTQASHAEANQQQTLKELESQVSC 1198
 QY 325 LQAQAEMLKEAERHQ-----RKQDPLLKAVE-SORMCELMKQQLTHLKKQ----- 370
 DB 1199 LEKFTSELKEAEQOKGNVLDLRKNKWAEMALAEACEBKLSRLTQAKSESEKQLHL 1258
 QY 371 -----LAL-----YTEKPEEF----- 381
 DB 1259 AEAQTKETLALLPGLSISAHQNYAEWLQEFKEKGSSELLKPPTEPSMDIVLKLREASE 1318
 QY 382 -QNTLSKSSEVTTTFQOHEKQTKVKKK-LEKETTMYRSRWESSNKALLEAEKTVROK 439
 DB 1319 TQNSLQAECDOYRTILAETEGMLKDLQKSVEEBEV-----WKAKVGAABEELHKSRTVK 1374
 QY 440 ELEGLOVKIQ-----RLEKLCRALQTERNDLNKRV---QDLSAGGQG 478
 DB 1375 HLEDIVELKLGHESSDQVREHTSHLEAELEKHAASAACQNYAKEVAGLRQLLESQS 1434
 QY 479 SUTDQGPERR-----PQPGQAQAPSSPRVTEAP 506
 DB 1435 QLDEAKSEAKQSDALALVRQQLSDMRSHVEDGVAGSPVAPPAEQDP 1492

RESULT 38

MYH6_MESAU STANDARD; PRT; 1939 AA.
 AC F13539; Q60562;
 DT 01-JAN-1990 (Rel. 13, Created)
 DT 15-DEC-1998 (Rel. 37, Last sequence update)
 DT 15-MAR-2004 (Rel. 43, Last annotation update)
 DE Myosin heavy chain, cardiac muscle alpha isoform (MyHC-alpha).
 GN MYH6.
 OS Mesocricetus auratus (Golden hamster).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Cricetinae;
 OC Mesocricetus.
 OX NCBI_TaxID=10036;

RN SEQUENCE FROM N.A.
 RC STRAIN=FIB; TISSUE=Liver;
 RX MEDLINE=95115033; PubMed=7815459;
 RA Wang R., Sole M.J., Cukerman E., Liew C.-C.;
 RT "Characterization and nucleotide sequence of the cardiac alpha-myosin
 RL heavy chain gene from Syrian hamster";
 RJ J. Mol. Cell. Cardiol. 26:1155-1165(1994).
 [2]
 RN SEQUENCE OF 1630-1939 FROM N.A.
 RX MEDLINE=86205859; PubMed=3458174;
 RA Liew C.-C., Jandreski M.A.;
 RT "Construction and characterization of the alpha form of a cardiac
 RL myosin heavy chain cDNA clone and its developmental expression in the
 RJ Syrian hamster";
 RL Proc. Natl. Acad. Sci. U.S.A. 83:3175-3179(1986).
 CC -!- FUNCTION: Muscle contraction.
 CC -!- SUBUNIT: Muscle myosin is a hexameric protein that consists of 2
 CC heavy chain subunits (MHC), 2 alkali light chain subunits (MLC)
 CC and 2 regulatory light chain subunits (MLC-2).
 CC -!- SUBCELLULAR LOCATION: Thick filaments of the myofibrils.
 CC -!- DOMAIN: The rodlike tail sequence is highly repetitive, showing
 CC cycles of a 28-residue repeat pattern composed of 4 heptapeptides,
 CC characteristic for alpha-helical coiled coils.
 CC -!- MISCELLANEOUS: Each myosin heavy chain can be split into 1 light
 CC meromyosin (LMM) and 1 heavy meromyosin (HMM). It can later be
 CC split further into 2 globular subfragments (S1) and 1 rod-shaped
 CC subfragment (S2).
 CC -!- MISCELLANEOUS: The cardiac alpha isoform is a 'fast' ATPase
 CC myosin, while the beta isoform is a 'slow' ATPase.
 CC -!- SIMILARITY: Contains 1 myosin-like globular head domain.
 CC -!- SIMILARITY: Contains 1 IQ domain.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See [http://www.isb-sib.ch/](http://www.isb-sib.ch/announce/)
 CC or send an email to license@isb-sib.ch).
 CC
 CC EMBL; L15351; AAB59701.1; -;
 CC EMBL; M12995; AAA37081.1; -;
 CC PIR; I48175; I48175.
 CC HSP; F08799; LMM.
 CC InterPro; IPR000048; IQ region.
 CC InterPro; IPR001609; myosin head.
 CC InterPro; IPR004009; Myosin N.
 CC InterPro; IPR002928; Myosin tail.
 CC Pfam; PF00612; IQ; 2.
 CC Pfam; PF00963; myosin head; 1.
 CC Pfam; PF02736; Myosin_N; 1.
 CC Pfam; PF01576; Myosin_tail; 1.
 CC PRINTS; PR00193; MYOSINHEAVY.
 CC ProDom; PD000355; myosin_head; 1.
 CC SMART; SM00015; IQ; 1.
 CC SMART; SM00242; MYSC; 1.
 CC PROSITE; PS50096; IQ; 1.
 CC
 CC Myosin; Muscle protein; Coiled coil; Thick filament; Actin-binding;
 KW ATP-binding; Methylation; Multigene family; Calmodulin-binding.
 FT DOMAIN 1 782 MYOSIN HEAD-LIKE.
 FT 783 812 IQ.
 FT DOMAIN 842 1939 COILED COIL (POTENTIAL).
 FT 1439 1443 POLY-ALA.
 FT DOMAIN 178 185 ATP (POTENTIAL).
 FT NP_BIND 657 679 ACTIN-BINDING.
 FT DOMAIN 759 773 ACTIN-BINDING.
 FT MOD_RES 129 129 METHYLATION (TRI-) (POTENTIAL).
 FT 1633 1633 Q -> L (IN REF. 2).
 FT 1651 1651 H -> Q (IN REF. 2).
 FT 1686 1687 EL -> DV (IN REF. 2).
 FT 1693 1693 V -> G (IN REF. 2).
 FT 1844 1844 K -> R (IN REF. 2).
 FT 1844 1844 K -> R (IN REF. 2).

FT CONFLICT 1879 1879 A -> T (IN REF. 2).
 FT CONFLICT 1885 1885 E -> Q (IN REF. 2).
 FT CONFLICT 1907 1907 E -> V (IN REF. 2).
 FT CONFLICT 1928 1928 D -> N (IN REF. 2).
 FT CONFLICT 1933 1933 QKM -> KR (IN REF. 2).
 SQ SEQUENCE 1939 AA; 223626 MM; DBC8297DFE83115A CRC64;
 Query Match 8.5%; Score 230.5; DB 1; Length 1939;
 Best Local Similarity 24.4%; Pred. No. 0.027;
 Matches 146; Conservative 80; Mismatches 197; Indels 175; Gaps 25;
 QY 12 EGAQRPSCAPAVEAGPGSGAARPKGAGQAQTAQSGALRDVSEELSRQLEDILSTYC 71
 DB 1037 EGSLEQEKVMDLE-----RAKKLEGLDNLVQSS--INDLEND-KLQLEKKKKKE 1086
 QY 72 VDNNGQPGEDCAQGEPAEPEDA-----EKSRTYVAR-----NGEPEPTPVVYGEKEP 119
 DB 1087 FTISQ-----QNSKLEDEQALALQKLLKKNQARIEELEEELEAERTAKVKEKL 1138
 QY 120 SKGDDPNTBIROS-DEVG-----DRHRRPOEKKKAGLCKETLLMQTLNTLTSTPEE 171
 DB 1139 SDLTRELEIEISERLEEAGGATSVQIEMNKYKRAEFQKMRDUEATL-----QHEA 1189
 QY 172 KLAALCKKY---AELEHRNSQKMLKQKQKQKQKQKQKQKQKQKQKQKQKQKQKQKQK 227
 DB 1190 TAAALEKXKXADSVAELEGEQIDNLQKQKLEKSEKSEKSEKSEKSEKSEKSEKSEK 1249
 QY 228 LCR-----ELQHRNSKKEGVQAR-----NSK 248
 DB 1250 VSTLEDAQYRVKLEESQSLNDFTQRAKLTQENGSLARQLEKEKALISOLTRGKLS 1309
 QY 249 -----EEBKREKVTSH-FOVTLNDIQIQMEQHNR-----NSK 281
 DB 1310 YTOQMEDLRQLEEEGKAKNALAHQSAHDCDLAREQYEEBEMEAKEALQKRVLSKANSE 1369
 QY 282 LR-----QENMELARLKKLIEQVELRBEHIDKY-----FKHK---DLQK 318
 DB 1370 VAQWRTKYETDAIQRTTELEERAKKLAQRIQDAEEAVEAVNAKCSLEKTKHLEQNEIED 1429
 QY 319 QLYDAKLOQAQEMLKABSEHGEKDFLLKEAVESQRMCELMKQOETHLKQQLALAYTEKF 378
 DB 1430 LAMVDERSNA---AAALQKQKQNFKILAE---WKQYEESEQSSESSQKEARSLSLELF 1484
 QY 379 EEFQNTLSKSSRYFTTFKQEMEXOTTKI-----KKLEKETTMYR 417
 DB 1485 -KLKNAVESELEHETFKRENKQLQBEISDLTEQLGEGGKNNVHELEKVRKQLEVKMELQ 1543
 QY 418 SRWESNKALLEMAKRYVDKKEGLQVKIQRLKLC-----RALQTERNDLNKRVQD 471
 DB 1544 SALEEA-EASLEHEBGLILR-AQLSPNQIKASIERKLAKEDEEMQAKRNHL--RVD 1597
 RESULT 39
 ID_AKA9_HUMAN STANDARD; PRT; 3911 AA.
 AC Q99996; O14869; O43355; Q94895; Q9UOH3; Q9UQ04; Q9Y6B8; Q9Y6Y2;
 DT 16-OCT-2001 (Rel. 40, Created)
 DT 16-OCT-2001 (Rel. 40, Last sequence update)
 DE 15-MAR-2004 (Rel. 43, Last annotation update)
 DE A-kinase anchor protein 9 (Protein kinase A anchoring protein 9)
 DE (PKA9) (A-kinase anchor protein 450 kDa) (AKAP 450) (A-kinase anchor
 DE protein 350 kDa) (AKAP 350) (hgAKAP 350) (AKAP 120 like protein)
 DE (Hyperion protein) (Yotiao protein) (Centrosome- and Golgi-localized
 DE PKM-associated protein) (CG-NAP).
 GN AKA9 OR AKA9450 OR AKA9350 OR KIAA0803.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 ON NCBTAXID=9606;
 RN [1]
 RP SEQUENCE FROM N.A. (ISOFORM 4).
 RC TISSUE=Brain;
 RX MEDLINE=98151389; PubMed=9482789;

RA Lin J.W., Wyszynski M., Madhavan R., Sealock R., Kim J.U., Sheng M.,
RT "Yotiao, a novel protein of neuromuscular junction and brain that
RT interacts with specific splice variants of NMDA receptor subunit
RT NR1.",
RL J. Neurosci. 18:2017-2027(1998).
RN [2]
RP SEQUENCE FROM N.A. (ISOFORM 2), AND VARIANT GLN-1347 INS.
RX MEDLINE=9921984; PubMed=10202149;
RA Witczak O., Skaalnegg B.S., Keryer G., Bornens M., Tasken K.,
RT Jahnson T., Oerstavik S.,
RA "Cloning and characterization of a cDNA encoding an A-kinase anchoring
RT protein located in the centrosome, AKAP450.",
RL EMBO J. 18:1858-1868(1999).
RN [3]
RP SEQUENCE FROM N.A. (ISOFORM 3).
RC TISSUE=Brain;
RX MEDLINE=99287934; PubMed=10358086;
RA Takahashi M., Shibata H., Shimakawa M., Miyamoto M., Mukai H., Ono Y.,
RT "Characterization of a novel giant scaffolding protein, CG-NAP, that
RT anchors multiple signaling enzymes to centrosome and the Golgi
RT apparatus.",
RL J. Biol. Chem. 274:17267-17274(1999).
RN [4]
RP SEQUENCE FROM N.A. (ISOFORM 1).
RA Kemmer W.A., Deiss S., Schwarz U.,
RT "Cloning of Hyperion.",
RL Submitted (AUG-1998) to the EMBL/GenBank/DBJ databases.
RN [5]
RP SEQUENCE OF 323-3911 FROM N.A. (ISOFORM 2).
RC TISSUE=Gastric parietal cell;
RX MEDLINE=99115654; PubMed=9915845;
RA Schmidt P.H., Dransfield D.T., Claudio J.O., Hawley R.G.,
RT Trotter K.W., Milgram S.L., Goldenring J.R.,
RA "AKAP350, a multiply spliced protein kinase A-anchoring protein
RT associated with centrosomes.",
RL J. Biol. Chem. 274:3055-3066(1999).
RN [6]
RP SEQUENCE OF 1802-3876 FROM N.A. (ISOFORM 5).
RC TISSUE=Lymphoblast;
RX Hinds K., Sutterer C., Becker M., Hawkins M.,
RA Submitted (JAN-1998) to the EMBL/GenBank/DBJ databases.
RN [7]
RP SEQUENCE OF 2157-3911 FROM N.A. (ISOFORM 6).
RC TISSUE=Lung;
RA Milgram S.L., Goldenring J.R., Schmidt P.H.,
RT "AKAP350: A multiply spliced family of proteins with centrosomal
RT association.",
RL Submitted (SEP-1998) to the EMBL/GenBank/DBJ databases.
RN [8]
RP SEQUENCE OF 2212-3911 FROM N.A. (ISOFORM 2/3).
RC TISSUE=Brain;
RX MEDLINE=99087487; PubMed=9872452;
RA Nagase T., Ishikawa K.-I., Suyama M., Kikuno R., Miyajima N.,
RT Tanaka A., Kotani H., Nomura N., Ohara O.,
RT "Prediction of the coding sequences of unidentified human genes. XI.
RT The complete sequences of 100 new cDNA clones from brain which code
RT for large proteins in vitro.",
RL DNA Res. 5:277-286(1998).
RN [9]
RP SEQUENCE OF 17-1800 FROM N.A.
RA Wu X., Graves T., Bradshaw H.,
RL Submitted (SEP-1998) to the EMBL/GenBank/DBJ databases.
RN [10]
RP FUNCTION: Binds to type II regulatory subunits of protein kinase
CC A. Scaffolding protein that assembles several protein kinases and
CC phosphatases on centrosome and Golgi apparatus where physiological
CC events can be regulated by phosphorylation state of protein
CC substrates. Isoform 4/Yotiao is associated with the N-methyl-D-
CC aspartate receptor and is specifically found in the neuromuscular
CC junction (NMJ) as well as in neuronal synapses explaining that its
CC role may be to organize postsynaptic specializations.
CC -! SUBUNIT: Interacts with the regulatory region of protein kinase N
CC (PKN). Protein phosphatase 2A (PP2A), protein phosphatase 1 (PP1)
CC and the immature non-phosphorylated form of PKC epsilon.
CC

CC -! SUBCELLULAR LOCATION: Centrosomal in many cell types and
CC cytoplasmic in parietal cells.
CC -! ALTERNATIVE PRODUCTS:
CC Event=Alternative splicing; Named isoforms=6;
CC Name=1;
CC IsoId=Q99996-1; Sequence=Displayed;
CC Name=2;
CC IsoId=Q99996-2; Sequence=VSP_004102, VSP_004107;
CC Name=3; Synonyms=CG-NAP;
CC IsoId=Q99996-3; Sequence=VSP_004102, VSP_004105, VSP_004107;
CC Name=4; Synonyms=Yotiao;
CC IsoId=Q99996-4; Sequence=VSP_004103, VSP_004104;
CC Name=5;
CC IsoId=Q99996-5; Sequence=VSP_004108;
CC Name=6; Synonyms=AKAP350;
CC IsoId=Q99996-6; Sequence=VSP_004106, VSP_004107, VSP_004109;
CC TISSUE SPECIFICITY: Widely expressed. Isoform 4/Yotiao is highly
CC expressed in skeletal muscle and in pancreas.
CC -! DOMAIN: RI-binding site, predicted to form an amphipathic helix,
CC could participate in protein-protein interactions with a
CC complementary surface on the R-subunit dimer.
CC -! CAUTION: Ref.6 sequence differs from that shown due to two
CC frameshifts in positions 3782 and 3811.
CC -! CAUTION: Ref.9 sequence differs from that shown due to four
CC frameshifts in positions 29, 1653, 1699 and 1735.
CC

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>;
CC or send an email to license@isb-sib.ch).
CC

EMBL;	AJ131693;	CAB40713.1;	--
EMBL;	AB019691;	BAA78718.1;	--
EMBL;	AJ010770;	CAA0361.1;	--
EMBL;	AF026245;	AAB86384.1;	--
EMBL;	AF083037;	AAD22767.1;	--
EMBL;	AC004013;	BAB96867.1;	ALT_FRAME
EMBL;	AF091711;	AAD39719.1;	--
EMBL;	AB018346;	BAA34523.1;	--
EMBL;	AC000066;	AAC60380.1;	ALT_FRAME
Genew;	HGNC:379;	AKAP9.	
MM;	604001;	--	
GO;	GO:0005813;	C:centrosome;	TAS.
GO;	GO:0005856;	C:cytoskeleton;	TAS.
GO;	GO:0004973;	F:N-methyl-D-aspartate receptor-associated pr. . . ;	TAS.
GO;	GO:0005515;	F:protein binding;	TAS.
GO;	GO:0007165;	F:signal transduction;	TAS.
GO;	GO:0007268;	P:synaptic transmission;	TAS.
GO;	GO:0006810;	P:transport;	TAS.
Coiled coil;	Alternative splicing;	Polymorphism.	
DOMAIN	2554	2567	PXA-RII SUBUNIT BINDING DOMAIN.
DOMAIN	164	934	COILED COIL (POTENTIAL).
DOMAIN	944	1022	COILED COIL (POTENTIAL).
DOMAIN	1100	1185	COILED COIL (POTENTIAL).
DOMAIN	1253	1280	COILED COIL (POTENTIAL).
DOMAIN	1336	1392	COILED COIL (POTENTIAL).
DOMAIN	1434	1459	COILED COIL (POTENTIAL).
DOMAIN	1585	1659	COILED COIL (POTENTIAL).
DOMAIN	1857	2455	COILED COIL (POTENTIAL).
DOMAIN	2544	2561	COILED COIL (POTENTIAL).
DOMAIN	2603	2776	COILED COIL (POTENTIAL).
DOMAIN	3065	3092	COILED COIL (POTENTIAL).
DOMAIN	3124	3470	COILED COIL (POTENTIAL).
DOMAIN	3587	3689	COILED COIL (POTENTIAL).
DOMAIN	3726	3730	POLY-LEU.
DOMAIN	203	292	GLN-RICH.
DOMAIN	321	1010	GLU-RICH.
DOMAIN	1846	2772	Missing (in isoform 2 and isoform 3).
VARSPPLIC	17	28	/FTId=VSP_004102.

```

Db 636 RURTOLLPSHEELSKEDLEIEHRIENIEKLDKNLGHYYQOQIDGLQENMSQIETWOF 695
QY 379 ERFONTLGNKSSVP-----TTFKQRMEXMTKKIKLEKETTMYRSRWSSNKAL 427
Db 696 EK-DNLIITKQNLIIIEIKLQDQOOSLVNSSEMTLQINELQKEIEILR----- 744
QY 428 LEMAEKTVRDKELGSLQVKIQLEKLCRALQTERNDINKRVQDLSA 474
Db 745 -QSEVEKGTLEQSVESLQKTELEK---QMKKENDLQKFAQLEA 787

RESULT 40
MYH4 HUMAN
ID MYH4 HUMAN STANDARD; PRT; 1939 AA.
AC Q9Y6Z3;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 15-MAR-2004 (Rel. 43, Last annotation update)
DE Myosin heavy chain, skeletal muscle, fetal (Myosin heavy chain IId)
DE (MyHC-IIb).
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxId=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Skeletal muscle;
RC MEDLINE=99318869; PubMed=10388558;
RT Weiss A., Schiaffino S., Leinwand L.A.;
RA "Comparative sequence analysis of the complete human sarcomeric myosin
RL heavy chain family: implications for functional diversity.";
RL J. Mol. Biol. 290:61-75(1999).
CC -!- FUNCTION: Muscle contraction.
CC -!- SUBUNIT: Muscle myosin is a hexameric protein that consists of 2
CC heavy chain subunits (MHC), 2 alkali light chain subunits (MLC)
CC and 2 regulatory light chain subunits (MLC-2).
CC -!- SUBCELLULAR LOCATION: Thick filaments of the myofibrils.
CC -!- DOMAIN: The rodlike tail sequence is highly repetitive, showing
CC cycles of a 28-residue repeat pattern composed of 4 heptapeptides,
CC characteristic for alpha-helical coiled coils.
CC -!- MISCELLANEOUS: Each myosin heavy chain can be split into 1 light
CC meromyosin (LMW) and 1 heavy meromyosin (HMM). It can later be
CC split further into 2 globular subfragments (S1) and 1 rod-shaped
CC subfragment (S2).
CC -!- SIMILARITY: Contains 1 myosin-like globular head domain.
CC -!- SIMILARITY: Contains 1 IQ domain.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; AF111783; AAD29949.1; -.
CC HSP; P13538; 2MYS.
CC Genew; HGNC:7574; MYH4.
CC MIM; 160742; -.
CC InterPro; IPR000048; IQ_region.
CC InterPro; IPR001609; myosin_head.
CC InterPro; IPR004009; Myosin_N.
CC InterPro; IPR002928; Myosin_tail.
CC Pfam; PF00612; IQ; 2.
CC Pfam; PF00063; myosin_head; 1.
CC Pfam; PF02736; Myosin_N; 1.
CC Pfam; PF01576; Myosin_tail; 1.
CC PRINTS; PR00033; MYOSINHEAVY.
CC ProDom; PD000355; myosin_head; 1.
CC SMART; SM00015; IQ; 1.
CC SMART; SM00242; MYSC; 1.

```

[illegible]


```
QY 286 -NWEAERUKKLEIYQVELREBHDKV-----FKHDLQOOL-----VDAKLQQAQ 329
DB 664 FOAHIEEKNVETLEKLDVQTELESLSSESVLKARHK-LEEELSVLKQDQDMKQBLE 722
QY 330 EMLKAEERHOREKDFLLKEAVESORMCB-----LMKQOETHLQO----- 369
DB 723 AKMDQKQHQQVDSIIKEHVSIVQTEKALQDQINQLELLKRDKXELKEHQAVENL 782
QY 370 -----QLALYTEKFEFQNTLSKSEVFTTFKQEMKTKIKKLEKTTWYRSRW 420
DB 783 EADIKSEGELOQASAKLDFVQSYQATHEQTAYEEQALQOQKLLDTERILL----- 838
QY 421 ESSNKALLEWAEKTVRDKEGLQVQIKORLEKLCALOTERNDLKRVQDLISAGQGSLL 480
DB 839 ----TKQVAEVAQKEVCTELDAHLQVQ----DLAQLEKQNSEKQVKSLTVYBSKL 892
QY 481 TDGSPER 487
DB 893 EDGNKEQ 899

RESULT 43
GOM4_MOUSE
ID GOA4_MOUSE STANDARD; PRT; 2238 AA.
AC Q91VW5; O70365; Q8CGH6;
DT 10-OCT-2003 (Rel. 42, Created)
DT 10-OCT-2003 (Rel. 42, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Golgi autoantigen, golgin subfamily A member 4 (tgolgin-1).
GN GOLGA4.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
SEQUENCE FROM N.A., TISSUE SPECIFICITY, AND SUBCELLULAR LOCATION.
RC STRAIN=BALB/c; TISSUE=Brain;
RX MEDLINE=22153218; PubMed=12162805;
RA Cowan D.A., Gay D., Bieler B.M., Zhao H., Yoshino A., Davis J.G.,
RA Tomayko M.M., Murali R., Greene M.I., Marks M.S.;
RT "Characterization of mouse Tgolgin-1 (golgin-245/trans-golgi p230/256
RT kd golgin) and its upregulation during oligodendrocyte development.";
RL DNA Cell Biol. 21:505-517(2002).
RN [2]
SEQUENCE OF 1-980 AND 991-2231 FROM N.A., AND VARIANTS ARG-61;
RP SER-280; GLU-293; SER-638; THR-819; THR-829 AND ALA-859.
RC STRAIN=C57BL/6, and Czech II; TISSUE=Brain, and Breast cancer;
RX MEDLINE=22388257; PubMed=12477932;
RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA Klausner R.D., Collins F.S., Wagner L., Shennan C.M., Schuler G.D.,
RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.B.,
RA Brownstein M.J., Udwin T.B., Toshiyuki S., Carninci P., Prange C.,
RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullaby S.J.,
RA Bonak S.A., McSwan P.J., McKernan K.J., Malek J.A., Gnarstine P.H.,
RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.G., Hulyk S.W.,
RA Vallalón D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA Fahey J., Helton E., Kettaman M., Madan A., Rodrigues S., Sanchez A.,
RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,
RA Butterfield Y.S.N., Krzywinski M.I., Skalska U., Smalius D.E.,
RA Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.;
RT "Generation and initial analysis of more than 15,000 full-length
RT human and mouse cDNA sequences.";
RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
CC -!- FUNCTION: May play a role in vesicular transport from the trans-
CC Golgi (By similarity).
CC -!- SUBCELLULAR LOCATION: Cytoplasmic, peripheral membrane protein
CC associated with the trans-Golgi network.
CC -!- TISSUE SPECIFICITY: Ubiquitous. Highly expressed in
```

```
CC oligodendrocyte precursors, particularly at a stage just prior to
CC myelination.
CC -!- DOMAIN: Extended rod-like protein with coiled-coil domains.
CC -!- SIMILARITY: Belongs to the golgin family.
CC -!- SIMILARITY: Contains 1 GRIP domain.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; AF051357; AAC05573.2; -
CC EMBL; BC007485; AAH07485.1; -
CC EMBL; BC037641; AAH37641.1; ALT_INIT.
CC EMBL; BC053000; AAH53000.1; -
CC PIR; T14265; T14265.
CC MGI; MGI:1859646; Golga4.
CC GO; GO:0005794; C:Golgi apparatus; IDA.
CC InterPro; IPR000237; GRIP domain.
CC InterPro; IPR003345; M_repeat.
CC Pfam; PF01465; GRIP; 1.
CC Pfam; PF02370; M; 12.
CC DR; MGI:1859646; Golga4.
CC DR; GO:0005794; C:Golgi apparatus; IDA.
CC InterPro; IPR000237; GRIP domain.
CC InterPro; IPR003345; M_repeat.
CC FT; PROSITE; PS50913; GRIP; 1.
KW Golgi stack; Coiled coil; Polymorphism.
FT DOMAIN 156 2161 COILED COIL (POTENTIAL).
FT DOMAIN 335 1011 GLU-RICH.
FT DOMAIN 1855 2107 GLU-RICH.
FT DOMAIN 2178 2225 GRIP.
FT VARIANT 61 61 T -> R (in strain Czech II).
FT VARIANT 280 280 G -> S (in strain Czech II).
FT VARIANT 293 293 G -> E (in strain Czech II).
FT VARIANT 638 638 G -> S (in strain Czech II).
FT VARIANT 819 819 K -> T (in strain Czech II).
FT VARIANT 829 829 A -> T (in strain Czech II).
FT VARIANT 859 859 T -> A (in strain Czech II).
FT VARIANT 2065 2065 R -> S (in strain C57BL/6).
FT SEQUENCE 2238 AA; 257562 MW; 494EA2C11F0165B CRC64;
QY Query Match 8.5%; Score 229.5; DB 1; Length 2238;
QY Best Local Similarity 22.1%; Pred. No. 0.034;
QY Matches 119; Conservative 107; Mismatches 200; Indels 113; Gaps 19;
QY 15 QERPQQAAPVAEAFGSGSQAPRKPEGAQAR-----TAQSGALRDVSE-----ELSR 61
DB 424 QKEKSERAAFELEKALST--AQKTEDAQRRKVMDEQMKAVEASBEERLRLQHELSR 481
QY 62 QLEDILSTVCYDNNQXGFGEDQAQGEPAEPEDAEKSRITYVARNQSPPEPTVYVYGEKPEPK 121
DB 482 VQEAASVAKKNSEQVAAALQKLHAEELASKEQELSRLEAREERELQEQMRITALEKSRSE 541
QY 122 GDPNTEIRIQSDVSG-----DRHRRPQKKKAGLQKEI-----TILMOTLN 164
DB 542 YLKLTOEKEQBSLALBELELQKKAILTSEENKLOEQEABAYRTRILEFTLSKSLQ 601
QY 165 TLSTPEEKLAALC-----KKYAELEERHNSQKMKLLQKQSQVQEKDHLRGHS 216
DB 602 ESKTQSEHLAVHLEAKKNKKNKELTALAEQHT---EVEGLQQQDQSLWTE----- 649
QY 217 KAVLARSKLESICRELQHRNLSKEGVQVQARBEERKQKVTSHFQVTLNDI-QLOMEQH 275
DB 650 -----ELQSL-----QQHQAAVEELREKYQOEKDALKEKESLQFAHIQDMNVEKTLKL 699
QY 276 NERNSKLQENNELAERLK---KLTEQVELREBHDKVFKHKDLQQLVDAKLQQAQOE-- 330
DB 700 DKQKQELSVSSELSEALFARDQLABELSVLRGDAKDK---KQALEAELEQRHH 752
QY 331 -----MLKAEERHOREKDFLLKEAVESORMCB-----QQLALYTEKPE 379
DB 753 QREVSISIQEQBLTVBRAEKALQD--ELSLGALLDERDEHLRERQARQVDLEAHLQKSA 810
```

QY 380 -EFQNTLSK-----SSEVFTTFKQEMKMYKKIKLEKTTMYRSRWSSNKALL 428
Db 811 GELQALAKLILLHSQSAAREQAGAYERQLAQMOCKVLDLETERKSL-----TKQV 963
QY 429 ENABEKTVEDKELEGVQIKRLEKLCRALQTERNDLNKRVQDLSAGGQGLTDSGPER 487
Db 864 EMETHKGHVCELDQAQRAQVQLER-----QRSELEKVRSLAQQLDSQLKNSTVEK 915

RESULT 44
MYHA HUMAN
ID MYHA HUMAN STANDARD; PRT; 1976 AA.
AC P35580;
DT 01-JUN-1994 (Rel. 29, Created)
DT 01-FEB-1996 (Rel. 33, Last sequence update)
DT 15-MAR-2004 (Rel. 43, Last annotation update)
DE Myosin heavy chain, nonmuscle type B (Cellular myosin heavy chain, type B) (Nonmuscle myosin heavy chain-B) (NMHC-B).
GN MYH10.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=96025307; PubMed=7499478;
RA Phillips C.L., Yamakawa K., Adelstein R.S.;
RT "Cloning of the cDNA encoding human nonmuscle myosin heavy chain-B and analysis of human tissues with isoform-specific antibodies.";
RL J. Muscle Res. Cell Motil. 16:379-389(1995).
RN [2]
RP SEQUENCE OF 63-722 FROM N.A.
RX MEDLINE=91316803; PubMed=1860190;
RA Simons M., Wang M., McBride O.W., Kawamoto S., Yamakawa K., Gdula D., Adelstein R.S., Weir L.;
RT "Human nonmuscle myosin heavy chains are encoded by two genes located on different chromosomes.";
RL Circ. Res. 69:530-539(1991).
CC -!- FUNCTION: Cellular myosin appears to play a role in cytokinesis, cell shape, and specialized functions such as secretion and capping.
CC -!- SUBUNIT: Myosin is a hexameric protein that consists of 2 heavy chain subunits (MHC), 2 alkali light chain subunits (MLC) and 2 regulatory light chain subunits (MLC-2).
CC -!- DOMAIN: The rodlike tail sequence is highly repetitive, showing cycles of a 28-residue repeat pattern composed of 4 heptapeptides, characteristic for alpha-helical coiled coils.
CC -!- SIMILARITY: Contains 1 myosin-like globular head domain.
CC -!- SIMILARITY: Contains 1 IQ domain.

This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

EMBL; M69181; AAA99177.1; -;
DR PIR; A59252; A59252.
DR HSGP; P10587; 1BR2.
DR Genew; HGNC:7568; MYH10.
DR MIM; 160776; -;
DR GO; GO:0016459; C:myosin; NAS.
DR GO; GO:0003779; F:actin binding; NAS.
DR GO; GO:0005524; F:ATP binding; NAS.
DR GO; GO:0000910; P:cytokinesis; NAS.
DR InterPro; IPR000048; IQ region.
DR InterPro; IPR001609; myosin head.
DR InterPro; IPR004009; myosin_N.
DR InterPro; IPR002928; myosin_tail.
DR InterPro; IPR002017; Spectrin.
DR Pfam; PF00612; IQ; 1.

Pfam; PF00063; myosin head; 1.
DR Pfam; PF02736; Myosin_N; 1.
DR Pfam; PF01576; Myosin_tail; 1.
DR PRINTS; PR00193; MYOSINHEAVY.
DR ProDom; PD000355; myosin_head; 1.
DR SMART; SM00015; IQ; 1.
DR SMART; SM00242; MYSC; 1.
DR PROSITE; PS00096; IQ; 1.
KW Myosin; ATP-binding; Calmodulin-binding; Actin-binding; Coiled coil;
KW Multigene family.
FT DOMAIN 1 785 MYOSIN HEAD-LIKE.
FT DOMAIN 786 815 IQ.
FT DOMAIN 845 1976 COILED COIL (POTENTIAL).
FT NP BIND 178 185 ATP (POTENTIAL).
SQ SEQUENCE 1976 AA; 228938 MW; B2BB87FF35EA124F CRC64;
Query Match 8 5%; Score 229; DB 1; Length 1976;
Best Local Similarity 24.2%; Pred No. 0.031;
Matches 93; Conservative 76; Mismatches 116; Indels 100; Gaps 16;
QY 160 MQLTNTLTPEEKLAALCKYAEILLEHNRNSQKMKLQKXOSQLVQSKDH----- 211
Db 841 VKPLAQVTRQEEELQAKDELLKVKQTKVEGELEEMERKHQQLLEKNILAEQLQAE 900
QY 212 -----RGEHSAVLARSK--LESCLRELQHRNSLKEEGVQARAREEK-----RKEVT 258
Db 901 ELFAEAEEMRAFLAAKQELSEILHDL-----SRVEEERNRQILQNEKKWQ 949
QY 259 SHFQVTLNDIQLOMEQHNRNSKLFQENMELAEKLIQVLELREHIDKVKFKHQLQQ 318
Db 950 AHIQ-----DLSEQLDEBEGARQLQLEKTAETAKIKKWESEILLLELDQNSKPKKELME 1005
QY 319 -----QLVD-----AKLQQAQ--MLKRAERHQRE-----KDFLLKSAV 351
Db 1006 DRIABCSQLAEEERKAKNLAKIRNKQVMISDLERLKEERKQLEKAKKLDGETT 1065
QY 352 ESQRMCELMKQETHLKOQLALYTEKFBFQNTLSKSEVFTTFKQEMKMYK-----KIT 407
Db 1066 DLQDQIAELQAIQIDELKLQLA---KKESELQALARGDDE--TLKNNALKVVRELAQIA 1121
QY 408 KLEKETTMYRSWESSNKALLEAEKTVRD---KELEGLOVKI----- 448
Db 1122 ELQED---PESEKASRNKA-----EKQRDLSELEALKTELEDTLDTAAQQELRTKR 1172
QY 449 -QLEKLCALQTERNDLNKRVQDL 472
Db 1173 EQVAELKKALEETKNNHEAQIQDM 1197

RESULT 45
LVA_DROME
ID LVA_DROME STANDARD; PRT; 2779 AA.
AC Q8MSS1; Q9W4N7;
DT 10-OCT-2003 (Rel. 42, Created)
DT 10-OCT-2003 (Rel. 42, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Lva lamp protein.
GN LVA OR CG6450.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Berkely;
RX MEDLINE=20196006; PubMed=10731132;
RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D., Ananides P.G., Scherer S.B., Li P.W., Hoskins R.A., Galie R.F., George R.A., Lewis S.B., Richards S., Ashburner M., Henderson S.N., Sutton G.G., Wortman J.R., Vandeil M.D., Zhang Q., Chen L.X., Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D., Wan K.H., Doyle C., Baxter C., Nelson C.R., Helt G., Miklos G.L.G.,

RA April J.F., Agbayani A., An H.-J., Andrews-Pfannkuch C., Baldwin D.,
RA Balow R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA Beeson K.X., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brotier P.,
RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA Cherry J.M., Cawley S., Dahike C., Davenport L.B., Davies P.,
RA de Pablo S., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA Dedson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,
RA Foster C., Gabriellian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA Glodek A., Gong P., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA Harris N.L., Harvey D.A., Helman T.J., Hernandez J.R., Houck J.,
RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lei Z.,
RA Lasco P., Lei Y., Levitsky A.A., Li J.H., Li Z., Liang Y., Lin X.,
RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzzy D.M., Nelson D.L.,
RA Nelson D.R., Nelson K.A., Nixon K.A., Nuskern D.R., Pacleb J.M.,
RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA Shue B.C., Sider-Kiamos I., Simpson M., Skupski K.P., Smith T.,
RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA Svrskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA Wang Z.-Y., Wasserman D.A., Weinstein G.M., Weissbach J.,
RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.,
RT "The genome sequence of *Drosophila melanogaster*."
RL Science 287:2185-2195(2000).
RN [2]
RP SEQUENCE OF 980-2779 FROM N.A.
RC STRAIN=Berkeley; TISSUE=Embryo;
RX MEDLINE=22426066; PubMed=12537569;
RA Stapleton M., Carlson J.W., Brokstein P., Yu C., Champe M.,
RA George R.A., Guarini H., Krommiller B., Pacleb J.M., Park S., Wan K.H.,
RA Rubin G.M., Gelinker S.B.,
RT "A *Drosophila* full-length cDNA resource."
RL Genome Biol. 3:RESEARCH080.1-RESEARCH080.8(2002).
RN [3]
RP FUNCTION, SUBUNIT, AND SUBCELLULAR LOCATION.
RX MEDLINE=20530668; PubMed=11076973;
RA Sisson J.C., Ventura R., Royou A., Sullivan W.,
RT "Lava lamp, a novel peripheral golgi protein, is required for
RT *Drosophila* melanogaster cellularization."
RL J. Cell Biol. 151:905-918(2000).
CC -!- FUNCTION: Lava and spectrin may form a Golgi-based scaffold that
CC mediates interaction of Golgi bodies with microtubules and
CC facilitates Golgi-derived membrane secretion required for the
CC formation of furrows during cellularization.
CC -!- SUBUNIT: Interacts with CLIP-190 and spectrin separately.
CC -!- SUBCELLULAR LOCATION: Lva-alpha-spectrin and Lva-CLIP-190
CC complexes are found at the Golgi.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; AR003429; AAF45910.1; -
CC EMBL; AY118638; AAM50007.1; ALT_INIT.
CC FlyBase; FBgn0029688; lva.
CC
CC Developmental protein; Golgi stacked; Coiled coil.
CC
CC DOMAIN 52 85 COILED COIL (POTENTIAL).
CC FT DOMAIN 141 175 COILED COIL (POTENTIAL).
CC FT DOMAIN 220 507 COILED COIL (POTENTIAL).
CC FT DOMAIN 559 716 COILED COIL (POTENTIAL).
CC FT DOMAIN 751 1733 COILED COIL (POTENTIAL).
CC FT

FT DOMAIN 1785 1863 COILED COIL (POTENTIAL).
FT DOMAIN 1941 2433 COILED COIL (POTENTIAL).
FT DOMAIN 2504 2544 COILED COIL (POTENTIAL).
FT DOMAIN 2600 2641 COILED COIL (POTENTIAL).
FT CONFLICT 1211 1211 E -> K (IN REF. 2).
FT CONFLICT 1424 1424 A -> S (IN REF. 2).
FT CONFLICT 1506 1506 E -> D (IN REF. 2).
FT CONFLICT 2034 2034 A -> V (IN REF. 2).
FT CONFLICT 2069 2069 Q -> H (IN REF. 2).
FT CONFLICT 2133 2133 Q -> E (IN REF. 2).
FT CONFLICT 2154 2154 Q -> E (IN REF. 2).
FT CONFLICT 2160 2160 E -> V (IN REF. 2).
FT CONFLICT 2200 2200 L -> P (IN REF. 2).
FT CONFLICT 2217 2217 A -> T (IN REF. 2).
FT CONFLICT 2271 2271 E -> D (IN REF. 2).
SQ SEQUENCE 2779 AA; 315897 MW; 1CB3965102018AE CRC64;
Query Match 8.5%; Score 229; DB 1; Length 2779;
Best Local Similarity 21.7%; Pred. No. 0.043;
Matches 127; Conservative 98; Mismatches 223; Indels 136; Gaps 20;
QY 5 GQPEAGSGAQRPSQAP-----AVRAGPGSSQAPKPEGAQARTASGA 51
DB 619 QRRPASBGDAQKQVHPQPSHVLTQTEBEDSSGETLSQLRRELFQERGE 678
QY 52 LRDVSEELS-----ROLF-DILSTYCYVDNNOGFGEDGAQGEPA 89
DB 679 VLDKLELSAENLQARLESSSLQLQOREREKOLISTSTSSNLS-----Q 727
QY 90 EPEDAESRTVARNGPETPVYVG--EKSPKGDPTBEIQCSDVGDHRRPQEK 147
DB 728 ELSNQRSSSVVATLDAGEGPNLFCKEKSLSKINSELAIRKANDQAKFNWVKLAK 787
QY 148 KAGLGKEITLLMQLNTLSTP-----EKLALCKYAELELHNSOKMKLLOKKQ 201
DB 788 EAKNCHTQLSELLHKVKEASTAVETVTVVAVTAPNGKALAEYQLNAQNAELKAVI 847
QY 202 SOLVQKDHARGH--SKANLA-----RSKLSLCRELORHNSLKEEY 244
DB 848 SKLQELDELRESYPETEAIPAIVGSDQREDEILOQSLQED-ARSLQAEQOIIEQV 906
QY 245 ---QRAREERERKEVTSHFQVTLNDIQLQMEQHNE-RNSKLQENNELAKLKEEQY 300
DB 907 DOIKELRQTEAQQLVARQSAEITQLQSEQPDQLNSKMSHEKQL-EQOTRIRREL 965
QY 301 ELREHIDKVPK-----HKLOQQLVDA-----KLQQAQEMLKE---AEERH 339
DB 966 EARAESLEGELSLIQLTVAEQKQOLIESVSESEHALNKMLELQSAQBELRLAKEDPD 1025
QY 340 QREKDFLLKEAVESQRMCKMQOET--HLKQQLALY-----TEKTEEFQV--- 383
DB 1026 QLRALRVSKSLVAQVQVELTSSQETVDALNQIQEYQGLEHAHKEBQFNRLREKLK 1085
QY 384 -----TLSSKSEVFTTFKQEMKMTKIKLKEKTTMYRSRWSNKKALLEM 430
DB 1086 YALNKKRTQDNADLEQKQVELTSQLOEQBELVKQKEEVERPVDNHRVQLQQQVSKL 1145
QY 431 AERKTV-----RDKELEGLOVKIQLREKLCLALQTERNDLN 466
DB 1146 NEDLKAKIHLNLRD-ALRQLKQIQEQQLQIERDAELQDAN 1188
RESULT 46
MYS_AEQIR
ID MYS_AEQIR STANDARD; PRT; 1938 AA.
AC P24733;
DT 01-MAR-1992 (Rel. 21, Created)
DT 01-MAR-1992 (Rel. 21, Last sequence update)
DT 15-MAR-2004 (Rel. 43, Last annotation update)
DE Myosin heavy chain, striated muscle.
OS Aequipecten irradians (Bay scallop).
OC Eukaryota; Metazoa; Mollusca; Bivalvia; Pteriomorpha; Pectinoida;
OC Pectinoidea; Pectinidae; Argopecten.

OX NCBI_TaxID=31199;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Adductor muscle;
RX MEDLINE=92011595; PubMed=1917970;
RA Nyitray L., Goodwin E.B., Szent-Gyorgyi A.G.;
RT "Complete primary structure of a scallop striated muscle myosin heavy
chain. Sequence comparison with other heavy chains reveals regions
that might be critical for regulation.";
RL J. Biol. Chem. 266:18469-18476 (1991).
RN [2]
RP SEQUENCE FROM N.A.
RC TISSUE=Adductor muscle;
RX MEDLINE=9108319; PubMed=2263488;
RA Nyitray L., Goodwin E.B., Szent-Gyorgyi A.G.;
RT "Nucleotide sequence of full length cDNA for a scallop striated
muscle myosin heavy chain.";
RL Nucleic Acids Res. 18:7158-7158 (1990).
RN [3]
RP X-RAY CRYSTALLOGRAPHY (2.8 ANGSTROMS) OF 777-836.
RX MEDLINE=9417332; PubMed=8127365;
RA Xie X., Harrison D.H., Schlichting I., Sweet R.M., Kalabokis V.N.,
Szent-Gyorgyi A.G., Cohen C.;
RT "Structure of the regulatory domain of scallop myosin at 2.8-A
resolution.";
RL Nature 368:306-312 (1994).
RN [4]
RP X-RAY CRYSTALLOGRAPHY (2.0 ANGSTROMS) OF 777-836.
RX MEDLINE=96419133; PubMed=8805510;
RA Houdusse A., Cohen C.;
RT "Structure of the regulatory domain of scallop myosin at 2-A
resolution: Implications for regulation.";
RL Structure 4:21-32 (1996).
CC -!- FUNCTION: Myosin is a protein that binds to F-actin and has ATPase
activity that is activated by F-actin.
CC -!- SUBUNIT: Muscle myosin is a hexameric protein that consists of 2
heavy chain subunits (MHC), 2 alkali light chain subunits (MLC)
and 2 regulatory light chain subunits (MLC-2).
CC -!- SUBCELLULAR LOCATION: Thick filaments of the myofibrils.
CC -!- SIMILARITY: Contains 1 myosin-like globular head domain.
CC -!- SIMILARITY: Contains 1 IQ domain.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
or send an email to license@isb-sib.ch).
CC -----
DR EMBL; X55714; CAA39247.1; -
DR PIR; A40997; A40997.
DR PDB; 1SCM; 30-APR-94.
DR PDB; 1WDC; 11-JUL-96.
DR PDB; 1B7T; 12-MAY-99.
DR PDB; 1KK7; 09-OCT-02.
DR PDB; 1KQW; 20-NOV-02.
DR PDB; 1KNO; 20-NOV-02.
DR PDB; 1L20; 30-OCT-02.
DR InterPro; IPR000048; IQ region.
DR InterPro; IPR001609; myosin head.
DR InterPro; IPR004009; myosin N.
DR InterPro; IPR002928; Myosin tail.
DR InterPro; IPR002017; Spectrin.
DR Pfam; PF00612; IQ; 2.
DR Pfam; PF00063; myosin_head; 1.
DR Pfam; PF02736; Myosin N; 1.
DR Pfam; PF01576; Myosin tail; 1.
DR PRINTS; PR00193; MYOSINHEAVY.
DR ProDom; PD000355; myosin_head; 1.
DR SMART; SM00015; IQ; 1.
DR SMART; SM00242; MYSC; 1.

DR PROSITE; PS50096; IQ; 1.
KW Myosin; Muscle protein; Coiled coil; Thick filament; Actin-binding;
KM ATP-binding; Calmodulin-binding; 3D-structure.
FT DOMAIN 1 777
FT DOMAIN 778 805
FT DOMAIN 806 1938
FT DOMAIN 836 1938
FT NP_BIND 176 183
FT HELIX 776 822
FT TURN 823 833
FT TURN 834 836
FT TURN 837 838
SQ SEQUENCE 1938 AA; 22821 MW; A5CCB4127D1A4896 CRC64;
Query Match 8.4%; Score 228; DB 1; Length 1938;
Best Local Similarity 22.4%; Pred. No. 0.033;
Matches 112; Conservative 95; Mismatches 197; Indels 95; Gaps 17;
QY 47 AQSGALRDVSEBLSROLEDI--LSYCVNDNQGGGAGEPAEDAEKSRVYAR 103
DB 1272 SQRSLQANSLSUTQLQEDAEHRVSVLSKEKLSQLEDAEDARSLEETARSKLQNEVR 1331
QY 104 NGEPEPTPV--VYGEKPSKGDPTNTEIRQSDEVGRDRHRRPQEKKAQGLKEITLLMQ 161
DB 1332 NMHADMDAIRQLEEEBQBSKSDVQRLSKANNEI-----QWRSKPESEGANRTBEL--- 1383
QY 162 TLNLTSTPEKLAALCKYABLLLEHNSQKQKMLLOKKSQVQSKDHLRGEHSAVLA 221
DB 1384 -----EDQKRLGLKLSAEQTTEAANAKCSALEKAKSRLOQE----- 1421
QY 222 RSKLESICRELQHRNRSKAE--EGVQRA-----REEBEKRKEVTSHPQVTL 265
DB 1422 ---LEDXSIEVDRAANAVNCKEKKQAFDKTTAEWQAKVNSLQSELENSKESRGYSAEL 1478
QY 266 NDITQLQNGHNRNSKLRQNMELAEKRLKLEIYVEL--REEH-IDKVFHKLQQLQVLD 322
DB 1479 YRIKASIEEYQDSIGALRRKNLADEIHDLTDQLSEGGSTHELDKARRRLEMEKEELQ 1538
QY 323 AKLQQAQEMLKEAEERHOR-----EKDPLLAEVES-----QRMCELMK--- 361
DB 1539 AALEEAGALEQBEAKVMRAQLFIATVNEIDKRIQEKKEEFNTNRNHQRALESQASL 1598
QY 362 QQTHLKLQALVTEKPEPQNTLSKSSVFTFFKQEMKQTKIKK-----LEKET 413
DB 1599 EBAKAGKADAMRIKKKLEQDINELEVALDASNRKAEKTKVKRYQQQIREMQTSIEEEQ 1658
QY 414 TMYRSWESSNKALLMAEKEK--TVRDKLEGLQVKIQRLEKLCALQTERNDLNKRVQDL 472
DB 1659 RQDREARESTN-----MAERCTMSEGEVEELRAALEQAEARAKASDNLADANDRVNEL 1713
QY 473 SAGGQGSLLTDSGPERRPEG 491
DB 1714 TS--QVS-SVQGGKRLKLEG 1729
RESULT 47
CEP2 HUMAN
ID CEP2 HUMAN STANDARD; PRT; 2442 AA.
AC Q9BV73; O14812; O60586; Q9H450;
DT 10-OCT-2003 (Rel. 42, Created)
DT 10-OCT-2003 (Rel. 42, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Centrosomal protein 2 (Centrosomal Nek2-associated protein 1) (C-NAP1)
DE (Centrosome protein 250) (Centrosome associated protein CEP250).
GN CEP2 OR CNAP1.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A. (ISOFORM 1), AND AUTOIMMUNE DISEASE.
RC TISSUE=Cervical carcinoma;
RX MEDLINE=98165428; PubMed=9506584;

RA	Mack G.J., Rees J., Sandblom O., Balczon R., Fritzler M.J., Rattner J.B.; "Autoantibodies to a group of centrosomal proteins in human autoimmune sera reactive with the centrosome."; <i>Arthritis Rheum.</i> 41:551-558(1998).	RT	localized to centrosomes, is complexed to protein phosphatase 1."; <i>Biochem. J.</i> 349:509-518(2000).
RA	[2]	RL	[6]
RA	SEQUENCE FROM N.A. (ISOFORMS 1 AND 2). INTERACTION WITH NEK2, AND SUBCELLULAR LOCATION DURING THE CELL CYCLE.	RN	PHOSPHORYLATION DURING CELL CYCLE.
RA	TISSUE=Placenta;	RX	Medline=22135747; PubMed=12140259;
RA	Fry A.M., Mayor T., Meraldi P., Stierhof Y.-D., Tanaka K., Nigg E.A.; "C-Nap1, a novel centrosomal coiled-coil protein and candidate substrate of the cell cycle-regulated protein kinase Nek2."; <i>J. Cell Biol.</i> 141:1563-1574(1998).	RA	Mayor T., Hacker U., Stierhof Y.-D., Nigg E.A.; "The mechanism regulating the dissociation of the centrosomal protein C-Nap1 from mitotic spindle poles."; <i>J. Cell Sci.</i> 115:3275-3284(2002).
RA	[3]	RT	FUNCTION: Probably plays an important role in centrosome cohesion during interphase.
RA	SEQUENCE FROM N.A.	CC	!- SUBUNIT: Monomer and homodimer (Probable). Forms a complex in vitro with both NEK2 kinase and the PP1CC catalytic subunit of protein phosphatase 1 (PP1).
RA	Medline=21638749; PubMed=11780052;	CC	!- SUBCELLULAR LOCATION: Component of the core centrosome. In interphase cells, it specifically associates with the proximal ends of both mother and daughter centrioles. Associates with the centrosome in interphase cells. In mitotic cells, it dissociates from the mitotic spindle poles. At the end of cell division, it reaccumulates at centrosomes.
RA	Deloukas P., Matthews L.H., Ashurst J., Burton J., Gilbert J.G.R., Jones M., Stavrides G., Almeida J.P., Babbage A.K., Bagguley C.L., Bailey J., Barlow K.F., Bates K.N., Beard L.M., Beare D.M., Beasley O.P., Bird C.P., Blakey S.E., Bridgeman A.M., Brown A.J., Buck D., Burrill W.D., Butler A.P., Carder C., Carter N.P., Chapman J.C., Clamp M., Clark G., Clark L.N., Clark S.Y., Clee C.M., Clegg S., Cobley V.S., Collier R.E., Connor R.E., Corby N.R., Coulson A., Coville G.J., Deadman R., Dhami P.D., Dunn M., Ellington A.G., Frankland J.A., Fraser A., French L., Garner P., Grafham D.V., Griffiths C., Griffiths M.N.D., Gilliam R., Hall R.E., Hammond S., Harley J.L., Heath P.D., Ho S., Holden J.L., Howden P.J., Huckle E., Hunt A.R., Hunt S.E., Jekosch K., Johnson C.M., Johnson D., Kay M.P., Kimberley A.M., King A., Knights A., Laird G.K., Lawlor S., Levasialho M.H., Leversha M.A., Lloyd C., Lloyd D.M., Lovell J.D., Marsh V.L., Martin S.L., McCommachie L.J., McKay K., McMurray A.A., Milne S.A., Mistry D., Moore M.J.F., Mullikin J.C., Nickerson T., Oliver K., Parker A., Patel R., Pearce T.A.V., Beck A.I., Phillimore B.J.C.T., Prathalingam S.R., Plumb R.W., Ramsay H., Rice C.M., Ross M.T., Scott C.E., Sehra H.K., Showken R., Sims S., Skuce C.D., Smith M.L., Soderlund C., Stewart C.A., Sultston J.B., Swann R.M., Symamore N., Taylor R., Tee L., Thomas D.W., Thorpe A., Tracey A., Symans A.C., Vaudin M., Wall M., Wallis J.M., Whitehead S.L., Whittaker P., Willey D.L., Williams L., Williams S.A., Wilming L., Wray P.W., Hubbard T., Durbin R.M., Bentley D.R., Beck S., Rogers J.; "The DNA sequence and comparative analysis of human chromosome 20."; <i>Nature</i> 414:865-871(2001).	CC	!- ALTERNATIVE PRODUCTS: Event=Alternative splicing; Named isoforms=3; Name=1; IsoId=Q9BV73-1; Sequence=Displayed;
RA	[4]	CC	Name=2; IsoId=Q9BV73-2; Sequence=VSP_007372;
RA	SEQUENCE FROM N.A. (ISOFORM 3).	CC	Name=3; IsoId=Q9BV73-3; Sequence=VSP_007371;
RA	TISSUE=Placenta;	CC	Note=No experimental confirmation available;
RA	Medline=22389257; PubMed=12477932;	CC	!- TISSUE SPECIFICITY: Ubiquitously and weakly expressed.
RA	Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G., Klausner R.D., Collins P.S., Wagner L., Shenmen C.M., Schuler G.D., Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K., Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F., Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L., Brownstein M.J., Udell T.B., Toshiyuki S., Carninci P., Scheetz T.E., Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C., Raba S., Loquellano N.A., Peters G.J., Abramson R.D., Mullaby S.J., Bosak S.A., McEwan P.J., McKernan K.J., Mallek J.A., Gunaratne P.H., Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W., Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A., Fahey J., Helton E., Kettman M., Madan A., Rodriguez S., Sanchez A., Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G., Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C., Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M., Butterfield V.S.N., Krywinski M.I., Skalska U., Smalios D.E., Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.; "Generation and initial analysis of more than 15,000 full-length human and mouse cDNA sequences."; <i>Proc. Natl. Acad. Sci. U.S.A.</i> 99:16899-16903(2002).	CC	!- DISASS: Antibodies against CEP2 are present in sera from patients with autoimmune diseases that developed autoantibodies against centrosomal proteins.
RA	[5]	CC	This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See http://www.isb-sib.ch/announce/ or send an email to license@isb-sib.ch).
RA	Phosphorylation, and interaction with NEK2 and PP1CA.	DR	EMBL; AF022655; AAC06349.1; -
RA	Medline=21060765; PubMed=10880350;	DR	EMBL; AF049105; AAC07988.1; -
RA	Helps N.R., Luo X., Barker H.M., Cohen P.T.W.; "NIMA-related kinase 2 (Nek2), a cell-cycle-regulated protein kinase	DR	EMBL; AL121586; CAB89415.1; -
RT		DR	EMBL; BC001433; AAH01433.1; -
RT		DR	Gene; HGNC:1859; CEP2.
RT		KW	Cell cycle; Coiled coil; Phosphorylation; Alternative splicing; Polymorphism.
RT		KW	DOMAIN 95 158 COILED COIL (POTENTIAL).
RT		FT	DOMAIN 244 352 COILED COIL (POTENTIAL).
RT		FT	DOMAIN 395 1172 COILED COIL (POTENTIAL).
RT		FT	DOMAIN 1243 2227 COILED COIL (POTENTIAL).
RT		FT	DOMAIN 2262 2376 COILED COIL (POTENTIAL).
RT		FT	DOMAIN 246 250 POLY-LEU.
RT		FT	DOMAIN 464 2171 GLN/LEU-RICH.
RT		FT	VARSPLIC 313 313 V -> F (in isoform 3).
RT		FT	VARSPLIC 314 2442 /FTId=VSP_007370.
RT		FT	VARSPLIC 863 918 /FTId=VSP_007371.
RT		FT	VARSPLIC 995 995 Missing (in isoform 2).
RT		FT	VARIANT 120 120 Q -> H (in dbSNP:2296403).
RT		FT	CONFLICT 136 136 L -> I (IN REF. 1).
RT		FT	CONFLICT 365 365 E -> A (IN REF. 1).
RT		FT	CONFLICT 365 365 H -> L (IN REF. 1).

FT CONFLICT 372 372 D -> E (IN REF. 1).
 FT CONFLICT 509 509 E -> D (IN REF. 2).
 FT CONFLICT 552 552 S -> I (IN REF. 1).
 FT CONFLICT 757 757 E -> A (IN REF. 1).
 FT CONFLICT 784 787 EVTK -> DEPO (IN REF. 1).
 FT CONFLICT 1153 1153 Q -> H (IN REF. 1).
 FT CONFLICT 1246 1246 H -> L (IN REF. 1).
 FT CONFLICT 1513 1513 L -> P (IN REF. 1).
 FT CONFLICT 2082 2082 Q -> L (IN REF. 1).
 FT CONFLICT 2345 2345 D -> N (IN REF. 1).
 SQ SEQUENCE 2442 AA; 281135 MW; BC28A36207B8272 CRC64;

 Query Match 8.4%; Score 228; DB 1; Length 2442;
 Best Local Similarity 23.1%; Pred. No. 0.042;
 Matches 122; Conservative 95; Mismatches 198; Indels 112; Gaps 19;

 QY 45 RTAGSGALRDVSELSRQLE-----DILSTYCVNNQGGFGEDGAQGEPAEPEDAFKSR 98
 DB 1433 REEVETLRGQIQLEQREMOXAKALELLSLDKRNQ-----EVDLQOE--QIQELEKCR 1486

 QY 99 TYVARNGEPTPVVYGEKPSKDPNTEIRQSDVYGDHRRPQKXKAGIKGKITL 158
 DB 1487 SVL-----EHLFVAVQERE--QKLTVQREIQRELEK--DRETQNVLEHQLLEKXQDM 1537

 QY 159 LMQTLNTLSTPEEKLAALCKKYABLLLEHNRNSQKMLKQKQSQLVQEK---DHLR--- 212
 DB 1538 IESORGQVQDLKKQVLTLECLALELENNHMECCQKLIKELQECQRETQVATHTLTL 1597

 QY 213 GEHKAVALRSK-----LES-----LCRLQHNRSLE-----EGVQARRE- 250
 DB 1598 EERSQELQAOSSQTHDLESHSTVLAELQEDQEVKSQRQIELOQKQHLFODLERD 1657

 QY 251 -----BEKKEVTSHPQVTLNDIQLQMEQHNRNSKLRQENNELAERLK--- 294
 DB 1658 QELMLQKRIQVLEQDQRTQKILEEDLEQIKLSRGRGLTQRLQMLQERAEKGPS 1717

 QY 295 -----KLI-----EQVELREHIDKVPKHD--LQQLV-----D 322
 DB 1718 KAQRSLSHMKLILRDKEKEVEQCEQHEHQLKDLQLEQQLQGLHHRKVGSETSLLSQRE 1777

 QY 323 AKLQAOEMLEKAEERHQREKDFILKEAVESORM-----CELMKQOETHKQQLALYT 375
 DB 1778 QEIVVQLOQAREEGELKQSQSLOSDEAQLAQDQLEALQEQQAQGEERVK 1837

 QY 376 EKFEFQNTLSKSESVFTTFQKEMKTKCKIKLEKETMYRSWESNKALLEMAEKT 435
 DB 1838 EKADALQALQEAHMTLKERHGLQDHKEQARLEELAVGRRVQALEEVLGDLRAER 1897

 QY 436 VRDKELEGLO-----VKIQLEKLCALQTE-----RNDLNRKVDLSA 474
 DB 1898 EQEKALLALQOQCAEQAEQHEVETRALQDSWMLQAOAVLEKRDQSEEA 1944

 RESULT 48
 ID US01_YEAST STANDARD; PRT; 1790 AA.
 AC P25386;
 DT 01-MAY-1992 (Rel. 22, Created)
 DT 01-MAY-1992 (Rel. 22, Last sequence update)
 DT 15-MAR-2004 (Rel. 43, Last annotation update)
 DE Intracellular protein transport protein US01.
 GN US01 OR INT1 OR YDL058W
 OS Saccharomyces cerevisiae (Baker's yeast).
 OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
 OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
 OX NCBI_TaxID=4932;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=X2180-1A;
 RX MEDLINE=911185402; PubMed=2010462;
 RA Nakajima H., Hirata A., Ogawa Y., Yonehara T., Yoda K.,
 RA Yamasaki M.;
 RT "A cytoskeleton-related gene, us01, is required for intracellular

protein transport in Saccharomyces cerevisiae.*;
 J. Cell Biol. 113:245-260(1991).
 [2]
 RN SEQUENCE OF 782-1790 FROM N.A.
 RA Hostetter M.K., Herman D.J., Bendel C.M., McClellan M., Tao N.,
 RA Kendrick K.E.;
 RL Submitted (FEB-1993) to the EMBL/GenBank/DBJ databases.
 RN [3]
 RN SEQUENCE OF 1-8 FROM N.A.
 RA Bai Y., Symington L.S.;
 RL Submitted (MAY-1996) to the EMBL/GenBank/DBJ databases.
 CC -!- FUNCTION: Required for protein transport from the ER to the Golgi
 complex.
 CC -!- SUBCELLULAR LOCATION: CYTOPLASMIC. ASSOCIATED WITH INTRACELLULAR
 MEMBRANES. PROBABLY PRESENT ON VESICLES OPERATIONAL BETWEEN THE
 ER AND THE GOLGI COMPLEX.
 CC -!- DOMAIN: The rodlike tail sequence is highly repetitive, composed
 of an heptapeptide repeat pattern characteristic of alpha-helical
 coiled coils. May form filamentous structures in the cell.
 CC -!- SIMILARITY: BELONGS TO THE VDP/USO1/YEL047C FAMILY.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation -
 the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 or send an email to license@sib-sib.ch).
 CC
 CC EMBL; X54378; CAA38253.1; -;
 CC EMBL; L03188; AAB00143.1; -;
 CC EMBL; U53669; AAB86659.1; -;
 CC GenOnline; 140300; -;
 CC SGD; S0002216; US01.
 CC InterPro; IPR008938; ARM.
 CC InterPro; IPR02017; Spectrin.
 CC InterPro; IPR006955; Usol_p115_C.
 CC InterPro; IPR006953; Usol_p115_head.
 CC Pfam; PF04871; Usol_p115_C; 1.
 CC Pfam; PF04869; Usol_p115_head; 1.
 CC Transprot; Protein transport; Golgi stack; Cytoskeleton; Coiled coil.
 KW DOMAIN 1 724 GLOBULAR HEAD.
 FT DOMAIN 725 1790
 FT DOMAIN 465 487
 FT DOMAIN 991 1790
 FT DOMAIN 1172 1785
 FT CONFLICT 847 847
 FT CONFLICT 924 924
 FT CONFLICT 1253 1253
 FT CONFLICT 1319 1319
 FT CONFLICT 1461 1461
 FT CONFLICT 1581 1581
 FT CONFLICT 1600 1600
 FT CONFLICT 1661 1661
 FT CONFLICT 1772 1772
 FT CONFLICT 1790 1790
 FT SEQUENCE 1790 AA; 206424 MW; 6CE2B21659FD4818 CRC64;

 Query Match 8.4%; Score 226.5; DB 1; Length 1790;
 Best Local Similarity 20.7%; Pred. No. 0.035;
 Matches 107; Conservative 99; Mismatches 173; Indels 139; Gaps 19;

 QY 52 LRDVSELSRQLEDILSTYCVNNQGGFGEDGAQGEPA-----EPEDAQS----- 97
 DB 802 VRDSDLEMT-QLRDVLKTKKEN-----QTALLEYKSTHKKQSDSIKLEKGL 848

 QY 98 RTVVARNGEPEPTPVVYGE-----KEPSKGDNTEIRQSDVYGDHRRPQKXKAG 151
 DB 849 ETLSQKKKAEADGINKMGKDLFALSREMQAVENCKNLQK-----EKDKSNVHAKETKS 903

 QY 152 LGKEITLTMQTLNTLSTPEEKLAALCKKYABLLLEHNRNSQKMLKQKQSQLVQEKDHL 211
 DB 904 LKEDIAAKITEIKALINENLEMKIQ-----NNLSKEKEHISK---ELVEYSRF 950

QY 135 VGDHRRPQKKKAGLGLK-RITLLMOTLNTLSTPEEKLAALCKYAELEHRSQK 193
Db 1207 QIDNLRVQKLEKSKBFLKIDLLSSMSVSKANLKI CRTLEDQSLRAGKVEE 1266
QY 194 MKLLQKQSOVLQVQKHLRGEHSHKAVLARSKLESLCRLQHRHSLKE--EGVORAREE 251
Db 1267 ---IQRSLSLQVQKLEKSKBFLKIDLLSSMSVSKANLKI CRTLEDQSLRAGKVEE 1323
QY 252 EKKKEVTSF-FQVTLNDILOMEQNER-----NSKLR-----OE 285
Db 1324 NKAKNALAHALOSSRHDCDLLEQYEEBQKAEIQLRALSANSEVAQWTKYETDAIOR 1383
QY 286 NMELARLKLQYELPEEHIDKY-----FRHGD-LQQLVDAKLQ-QAOEMLKEAE 336
Db 1384 TRELEAKKLAQRLQDSEQVEAVNAKASLETKQRLQGEVEDLMVDVERANSLAAL 1443
QY 337 BRHQEKDPLLEKAEVSRMCLMKQ--ETHLKQALYTKFEEFQNTLSKSEVETTF 395
Db 1444 DKQKQNF---KVLAEWTKCBESQAEASLKSRSLSLTF-KLKNAYEALDQLETV 1499
QY 396 K-----QEMKQTKKI-----KKLEKTTMYRGRWSSNKAL-----427
Db 1500 KRENKLEOFIADLQEAENGKTIHELEKSKQIELEKADIQALBEAEAALEHEBAKI 1559
QY 428 ---LEMAEKT-----VRDEKLEGLQVKIOR-LEKLCRALQTERNDLNKRVODLSAG 475
Db 1560 LAIQLELTQVKEIDSKIAEKDEBIEQLKBNVQRTVETVQSQALDAEVSREAR-LKKX 1618
QY 476 GQGLSTD 482
Db 1619 MEGDLNE 1625

RESULT 50
CENP_HUMAN STANDARD; PRT; 3210 AA.
AC F49454; Q13171; Q13246;
DT 01-FEB-1996 (Rel. 33, Created)
DT 01-FEB-1996 (Rel. 33, Last sequence update)
DT 15-MAR-2004 (Rel. 43, Last annotation update)
DE CENP-F kinetochore protein (Centromere protein F) (Mitosis) (AH
D3 antigen).
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID:9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Breast carcinoma;
RX MEDLINE=95348175; PubMed=7542657;
RA Liao H., Winkfein R.J., Mack G., Rattner J.B., Yen T.J.;
RT "CENP-F is a protein of the nuclear matrix that assembles onto
kinetochores at late G2 and is rapidly degraded after mitosis."
RL J. Cell Biol. 130:507-518(1995).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=95379848; PubMed=7651420;
RA Zhu X., Mancini M.A., Chang K.-H., Liu C.-Y., Chen C.-F., Shan B.,
Jones D., Yang-Feng T.L., Lee W.-H.;
RT "Characterization of a novel 350-kilodalton nuclear phosphoprotein
that is specifically involved in mitotic-phase progression."
RL Mol. Cell. Biol. 15:5017-5029(1995).
RN [3]
RP SEQUENCE OF 2194-3210 FROM N.A.
RX MEDLINE=95336446; PubMed=7612011;
RA Li Q., Ke Y., Kapp J.A., Fertig N., Medger T.A. Jr., Joshi H.C.;
RT "A novel cell-cycle-dependent 350-kDa nuclear protein: C-terminal
domain sufficient for nuclear localization."
RL Biochem. Biophys. Res. Commun. 212:220-228(1995).
RN [4]
RP CHARACTERIZATION.

RX MEDLINE=95370296; PubMed=7642639;
RA Zhu X., Chang K.-H., He D., Mancini M.A., Brinkley W.R., Lee W.-H.;
RT "The C terminus of mitotin is essential for its nuclear localization,
centromere/kinetochore targeting, and dimerization."
RL J. Biol. Chem. 270:19545-19550(1995).
RN [5]
RP CHARACTERIZATION.
RX MEDLINE=98437347; PubMed=9763420;
RA Chan K.T., Schaar B.T., Yen T.J.;
RT "Characterization of the kinetochore binding domain of CENP-E reveals
interactions with the kinetochore proteins CENP-F and HUBB1."
RL J. Cell Biol. 143:49-63(1998).
RN [6]
RP FARNESYLATION.
RX MEDLINE=20459117; PubMed=10852915;
RA Ashar H.R., James L., Gray K., Carr D., Black S., Armstrong L.,
Bishop W.R., Kirschmeier P.;
RT "Farnesyl transferase inhibitors block the farnesylation of CENP-E
and CENP-F and alter the association of CENP-E with the
microtubules."
RL J. Biol. Chem. 275:30451-30457(2000).
CC -!- FUNCTION: Probably required for kinetochore function, involved in
chromosome segregation during mitosis. Interacts with
retinoblastoma protein (RB), CENP-E and HUBB1.
CC -!- SUBUNIT: Homo- or heterodimer.
CC -!- SUBCELLULAR LOCATION: Nuclear matrix (but not in the nucleolus),
reorganization to the kinetochore/centromere (coronal surface of
the outer plate) and the spindle during mitosis.
CC -!- DEVELOPMENTAL STAGE: Gradually accumulates during the cell cycle.
CC -!- PTM: Hyperphosphorylated during mitosis.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
or send an email to license@sib-sib.ch).
CC EMBL; U19769; AAA82889.1; -
CC XBL; U30872; AAA82935.1; -
CC EMBL; U25725; AAA86889.1; -
CC PIR; PC4035; PC4035.
CC Genew; HGNC:1857; CENPF.
CC GK; P49454; -
CC MIM; 600236; -
CC GO; GO:0005699; C:kinetochore; TAS.
CC GO; GO:0005634; C:nucleus; TAS.
CC GO; GO:0005819; C:spindle; TAS.
CC GO; GO:0000067; P:DNA replication and chromosome cycle; TAS.
CC GO; GO:0007088; P:regulation of mitosis; TAS.
CC Chromosomal protein; Nuclear protein; Centromere; Coiled coil;
KW Mitosis; Phosphorylation; Antigen; Cell cycle; Repeat; Polymorphism;
KW Lipoprotein; Prenylation.
FT DOMAIN 14 197
FT COILED COIL (POTENTIAL).
FT DOMAIN 273 769
FT COILED COIL (POTENTIAL).
FT DOMAIN 823 1328
FT COILED COIL (POTENTIAL).
FT DOMAIN 1642 1746
FT COILED COIL (POTENTIAL).
FT DOMAIN 1862 2987
FT COILED COIL (POTENTIAL).
FT DOMAIN 2207 2568
FT COILED COIL (POTENTIAL).
FT REPEAT 2207 2386
FT REPEAT 2389 2568
FT DOMAIN 3015 3052
FT LIPID 3207 3207
FT VARIANT 3202 3202
FT /FTIG=VAR 014839.
FT T -> A (IN REF. 2).
FT L -> Q (IN REF. 2).
FT G -> D (IN REF. 2).
FT MISSING (IN REF. 2).
FT MISSING (IN REF. 2).
FT V -> A (IN REF. 2).
FT V -> L (IN REF. 2).

FT	CONFLICT	2242	2243	ER -> DG (IN REF. 3).
FT	CONFLICT	2335		L -> Q (IN REF. 3).
FT	CONFLICT	2492	2493	D -> N (IN REF. 2).
FT	CONFLICT	2545	2561	ELNVRALAHNDQACK -> SSMREWOPCMTWKPVV (IN REF. 3).
FT	SEQUENCE	3210 AA;	367589 MW;	11D83324960E4334 CRC64;
	Query Match	8.4%;	Score 226.5;	DB 1; Length 3210;
	Best local Similarity	22.4%;	Pred. No. 0.062;	
	Matches 102;	Conservative	91;	Mismatches 191; Indels 71; Gaps 14;
Qy	31	GSSQAPKPEGAQARTAQSGALRDVSELSQLEDIUSTYCVDNNGGFGEDGAQGEPAE	90	
Db	272	GNSSPHLLDQLK--AQQLNRKINLELR	305	---QGH--E 305
Qy	91	PDAEKSTYYVARNGEDEPTVWYCEKPSKGDNTBETIROSDVGDGRHREBPQKKAK	150	
Db	306	KEMKGQNVKFOELQLEKAKVELLEKRVUNKRDELUVRTTAQDQASTKYTALEQKLK	365	
Qy	151	GIGKEITILLMTLNTL-STPEKLAALCKYAELEEHNRNSQOMKLLQK	205	---OSQIV 205
Db	366	KLTEDLSQORNAESARCSLEQKIK--EKEKFOEELSROQRSFOTLDQECIQMKARLT	422	
Qy	206	QEKDHLRGEHKAVLARKLSLCRELOHNRSLKEBGVQARABEEKEKXVTSHPVTL	265	
Db	423	QELQAQKNHNVLAELDKLTSVKQOLENNLEEFKQ--KLCRABQAFQAS	474	---QJKE 474
Qy	266	NDIQLOMBOHNRNSKLROENMELA-----ERLKLIEQYELREBEHDKVFKHDLQOQ	319	
Db	475	NELRSMEEMKKNLLKSHSQKAREVCHLEABLKNIKCLNQSONFAEMKAKNTSQE	534	
Qy	320	LVDAKLQAQOEMLKABERHQBKDFLKEAVESQRMK--ELMKQOETHLKOQLALYIEK	377	
Db	535	TW--LRDLQEKINOENSLETKLKLAVADLEKORDCSQDLLKKREHHI	581	-----581
Qy	378	PEEFONTLSKSSEVFTTPKQEMKMTKKIKLEKETIMYRSRWSNSKALLEMASEKTVR	437	
Db	582	-EQLNDKLSKTEKESKALLSALELKKEVEELKEKTLF-SCKWSENEKLL	633	-----TQM 633
Qy	438	DKEGLEGVKIQRLEKLCRALQTERNDLNKRVQDL	472	
Db	634	ESEKENIASKINHLETCLTKTOQIKGHEVNERVTL	668	

Search completed: June 7, 2004, 14:38:07
Job time : 36 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: June 7, 2004, 14:16:35 ; Search time 134 Seconds
(without alignments)
1247.945 Million cell updates/sec

Title: US-10-023-529-8

Perfect score: 2702
Sequence: 1 KSPQPEAGPEGAQRPSQ.....APSTEASGQTGPQEPSARA 530

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 1017041 seqs, 315518202 residues

Total number of hits satisfying chosen parameters: 1017041

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 150 summaries

Database :

SPTREML 25: *
1: sp_archaea: *
2: sp_bacteria: *
3: sp_fungi: *
4: sp_human: *
5: sp_invertebrate: *
6: sp_mammal: *
7: sp_mhc: *
8: sp_organelle: *
9: sp_phage: *
10: sp_plant: *
11: sp_rodent: *
12: sp_virus: *
13: sp_vertebrate: *
14: sp_unclassified: *
15: sp_rvirus: *
16: sp_bacteriap: *
17: sp_archaeap: *

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Query	Score	Match	Length	DB ID	Description
1	2693	99.7	546	4	Q86YW3	Q86YW3 homo sapien
2	2689	99.5	546	4	Q86T86	Q86T86 homo sapien
3	2688	99.5	546	4	Q86T54	Q86T54 homo sapien
4	2677	99.1	546	4	Q86T85	Q86T85 homo sapien
5	1291	47.8	524	11	Q8BHN1	Q8BHN1 mus musculus
6	1283.5	47.5	528	4	Q9NUQ3	Q9NUQ3 homo sapien
7	1266.5	46.9	718	4	Q86T52	Q86T52 homo sapien
8	1265.5	46.8	715	4	Q8N313	Q8N313 homo sapien
9	1233	45.6	676	13	Q91969	Q91969 gallus gall
10	1223.5	45.5	684	11	Q8VBT1	Q8VBT1 mus musculus
11	1194	44.2	463	11	Q8BUK2	Q8BUK2 mus musculus
12	1008	37.3	505	4	Q8N3S2	Q8N3S2 homo sapien
13	646	23.9	241	11	Q8BP11	Q8BP11 mus musculus
14	598	22.1	116	4	Q86Y86	Q86Y86 homo sapien
15	597.5	22.1	515	5	Q9VBL3	Q9VBL3 drosophila
16	503	18.6	264	5	Q86LQ1	Q86LQ1 branchiostoma

17	462.5	17.1	436	10	Q84VE3	Q84ve3 oryza sativ
18	459	17.0	241	11	Q8BMJ7	Q8bmj7 mus musculus
19	450	16.7	404	10	Q8RDX7	Q8rxd7 arabidopsis
20	442.5	16.4	186	4	Q9P0X1	Q9p0x1 homo sapien
21	392.5	14.5	150	11	Q8C5K1	Q8c5k1 mus musculus
22	372.5	13.8	335	5	Q22666	Q22666 caenorhabdi
23	369.5	13.7	181	4	Q9BZA4	Q9bza4 homo sapien
24	332	12.3	372	10	Q9FGD8	Q9fgd8 arabidopsis
25	272	10.1	1168	5	Q9VYU0	Q9vyu0 drosophila
26	269	10.0	992	4	Q9NTH6	Q9nth6 homo sapien
27	266.5	9.9	900	3	Q9P3P5	Q9p3p5 neurospora
28	266.5	9.9	1455	4	Q9UPV0	Q9upv0 homo sapien
29	262	9.7	1937	6	Q9TV62	Q9tv62 sus scrofa
30	260	9.6	1941	13	Q9DGM4	Q9dgm4 gallus gall
31	259	9.6	1944	13	Q9DGM5	Q9dgm5 gallus gall
32	257	9.5	1943	13	Q8JG72	Q8jg72 gallus gall
33	256	9.5	1939	13	Q9PTY2	Q9pty2 gallus gall
34	254.5	9.4	609	17	Q8TXA4	Q8txa4 methanopyru
35	253	9.4	1206	11	Q08815	Q08815 rattus norv
36	253	9.4	1940	13	Q8AY28	Q8ay28 gallus gall
37	249.5	9.2	1233	11	Q54988	Q54988 mus musculus
38	249.5	9.2	2139	5	Q07569	Q07569 entamoeba h
39	249	9.2	610	11	Q9CS72	Q9cs72 mus musculus
40	249	9.2	1003	12	Q91LX9	Q91lx9 kaposi's sa
41	248.5	9.2	1005	4	Q9P216	Q9p216 homo sapien
42	247	9.1	1001	11	Q88664	Q88664 rattus norv
43	246.5	9.1	2055	5	Q8T5C7	Q8t5c7 plasmodium
44	246.5	9.1	2055	5	Q8IHP3	Q8ihp3 plasmodium
45	244	9.0	1202	11	Q9WU41	Q9wu41 mus musculus
46	244	9.0	1212	11	Q8K4T4	Q8k4t4 rattus norv
47	244	9.0	1307	11	Q80U65	Q80u65 mus musculus
48	244	9.0	1391	11	Q922J3	Q922j3 mus musculus
49	243.5	9.0	728	4	Q60561	Q60561 homo sapien
50	243.5	9.0	1001	4	Q9H2K7	Q9h2k7 homo sapien
51	243.5	9.0	1598	11	Q922D2	Q922d2 mus musculus
52	242	9.0	1937	6	Q8MJV1	Q8mjv1 equus caball
53	242	9.0	1938	6	Q8GUP9	Q8gup9 oryctolagus
54	242	9.0	1939	6	Q9TV63	Q9tv63 sus scrofa
55	242	9.0	1939	6	Q9TV61	Q9tv61 sus scrofa
56	241.5	8.9	1089	12	Q40947	Q40947 kaposi's sa
57	241.5	8.9	1508	5	Q86J39	Q86j39 dictyosteli
58	241.5	8.9	1364	5	Q8SNQ7	Q8snq7 loligo peal
59	241	8.9	1001	13	Q7Z100	Q7zyj0 xenopus lae
60	240.5	8.9	2033	10	Q7XKH4	Q7xeh4 oryza sativ
61	240	8.9	879	13	Q9YHD8	Q9yhd8 rana catesb
62	240	8.9	1931	13	Q91973	Q91973 coturnix co
63	240	8.9	1938	6	Q8MJV0	Q8mjv0 equus caball
64	239.5	8.9	901	11	Q7TN77	Q7tn77 mus musculus
65	239.5	8.9	1197	3	Q9C3Y7	Q9c3y7 candida alb
66	239.5	8.9	1201	11	Q8CIA8	Q8cia8 mus musculus
67	239.5	8.9	1413	11	Q8CGH3	Q8cgh3 mus musculus
68	239	8.8	1057	5	Q8MRS3	Q8mrs3 drosophila
69	239	8.8	1938	6	Q9BE40	Q9be40 bos taurus
70	239	8.8	1940	6	Q9BE41	Q9be41 bos taurus
71	239	8.8	2115	4	Q14980	Q14980 homo sapien
72	238.5	8.8	1935	6	Q9BE39	Q9be39 bos taurus
73	238	8.8	892	11	Q63939	Q63939 rattus sp.
74	238	8.8	1036	12	Q9DUM3	Q9dum3 kaposi's sa
75	237.5	8.8	1177	4	Q86TC3	Q86tc3 homo sapien
76	237.5	8.8	1213	4	Q727B0	Q727b0 homo sapien
77	237	8.8	631	13	Q7Z219	Q7zz19 brachydanio
78	237	8.8	915	5	Q86SR0	Q86sr0 dictyosteli
79	236	8.7	970	5	Q9N8Y9	Q9n8y9 trypanosoma
80	236	8.7	1935	6	Q8MJU9	Q8mjju9 equus caball
81	236	8.7	1935	11	Q91Z83	Q91z83 mus musculus
82	236	8.7	1941	13	Q8UWA0	Q8uwa0 gallus gall
83	235.5	8.7	1437	11	Q80XQ1	Q80xq1 mus musculus
84	235	8.7	728	11	Q8K454	Q8k454 mus musculus
85	235	8.7	1229	5	Q9NJ22	Q9nj22 aequipecten
86	235	8.7	1243	5	Q9NJ21	Q9nj21 aequipecten
87	235	8.7	1253	5	Q9NU20	Q9nu20 aequipecten
88	234.5	8.7	1951	5	Q17042	Q17042 aequipecten
89	234.5	8.7	650	4	Q8N9W4	Q8n9w4 homo sapien

RA Ansoorge W., Krieger S., Mewes H.W., Weil B., Amid C., Fobo G., Han M.,
RA Osanger A., Wiemann S.,
RL Submitted (APR-2003) to the EMBL/GenBank/DBJ databases.
DR EMBL; AL832636; CAD89951.1; -
KW Hypothetical protein.
SQ SEQUENCE 546 AA; 61892 MW; FEBCAD4F753F7FF1 CRC64;

Query Match 99.5%; Score 2689; DB 4; Length 546;
Best Local Similarity 99.6%; Pred. No. 1.6e-120;
Matches 528; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 KSSPGQPEAGPEGAQERPSQAAPAVEAGPGSSQAAPRKEGAQARTAGSALRDVSEELS 60
DB 17 KSSPGQPEAGPEGAQERPSQAAPAVEAGPGSSQAAPRKEGAQARTAGSALRDVSEELS 76
QY 61 RQLEDILSTYCVNNQGGPGEAGQEPAPAEPAEDAKSRITYVARNGEPPTPVVYGEKPS 120
DB 77 RQLEDILSTYCVNNQGGPGEAGQEPAPAEPAEDAKSRITYVARNGEPPTPVVYGEKPS 136
QY 121 KGPNTTEIRQSDVGDGRDHRPQEKKAKGLGKGIITLLMOTLNTLSTPEEKLAALCKKY 180
DB 137 KGPNTTEIRQSDVGDGRDHRPQEKKAKGLGKGIITLLMOTLNTLSTPEEKLAALCKKY 196
QY 181 AELEEHRSQKMKLLQKQSQVQKDLHGEHSKAVLARSKLESICRELQRNRSIK 240
DB 197 AELEEHRSQKMKLLQKQSQVQKDLHGEHSKAVLARSKLESICRELQRNRSIK 256
QY 241 EGVQARBEERKREKVTSHFQVTLNDIOLQMEQHNERNKSLRQNMELAEELKLIROY 300
DB 257 EGVQARBEERKREKVTSHFQVTLNDIOLQMEQHNERNKSLRQNMELAEELKLIROY 316
QY 301 ELREEHIDKVFHKDLQKQSQVQKDLHGEHSKAVLARSKLESICRELQRNRSIK 360
DB 317 ELREEHIDKVFHKDLQKQSQVQKDLHGEHSKAVLARSKLESICRELQRNRSIK 376
QY 361 KQETHLKOALALYTEKPEEFQNTLSKSEVFTTTPQEMEKMTKKIKKLEKTTMYRSRW 420
DB 377 KQETHLKOALALYTEKPEEFQNTLSKSEVFTTTPQEMEKMTKKIKKLEKTTMYRSRW 436
QY 421 ESSNKALLEMAEKTVDKELSGLOVKIQRLEKLCALQTERNDLNKRVQDLSAGGQSL 480
DB 437 ESSNKALLEMAEKTVDKELSGLOVKIQRLEKLCALQTERNDLNKRVQDLSAGGQSL 496
QY 481 TDSGPRRPEGGAQAPSPRVTEAPCPVGPAPSTASGQTGPQPTTSARA 530
DB 497 TDSGPRRPEGGAQAPSPRVTEAPCPVGPAPSTASGQTGPQPTTSARA 546

RESULT 3

Q86T54 ID Q86T54 PRELIMINARY; PRT; 546 AA.
AC Q86T54;
DT 01-JUN-2003 (TREMELrel. 24, Created)
DT 01-JUN-2003 (TREMELrel. 24, Last sequence update)
DE Hypothetical protein DKZP451K215.
GN DKZP451K215.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Human skeletal muscle;
RA Koehrer K., Beyer A., Mewes H.W., Weil B., Amid C., Osanger A.,
RA Fobo G., Han M., Wiemann S.,
RL Submitted (MAY-2003) to the EMBL/GenBank/DBJ databases.
DR EMBL; AL832338; CAD91138.1; -
KW Hypothetical protein.
SQ SEQUENCE 546 AA; 61861 MW; 728D674F78897DF6 CRC64;

Query Match 99.5%; Score 2688; DB 4; Length 546;
Best Local Similarity 99.6%; Pred. No. 1.8e-120;

Matches 528; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 KSSPGQPEAGPEGAQERPSQAAPAVEAGPGSSQAAPRKEGAQARTAGSALRDVSEELS 60
DB 17 KSSPGQPEAGPEGAQERPSQAAPAVEAGPGSSQAAPRKEGAQARTAGSALRDVSEELS 76
QY 61 RQLEDILSTYCVNNQGGPGEAGQEPAPAEPAEDAKSRITYVARNGEPPTPVVYGEKPS 120
DB 77 RQLEDILSTYCVNNQGGPGEAGQEPAPAEPAEDAKSRITYVARNGEPPTPVVYGEKPS 136
QY 121 KGPNTTEIRQSDVGDGRDHRPQEKKAKGLGKGIITLLMOTLNTLSTPEEKLAALCKKY 180
DB 137 KGPNTTEIRQSDVGDGRDHRPQEKKAKGLGKGIITLLMOTLNTLSTPEEKLAALCKKY 196
QY 181 AELEEHRSQKMKLLQKQSQVQKDLHGEHSKAVLARSKLESICRELQRNRSIK 240
DB 197 AELEEHRSQKMKLLQKQSQVQKDLHGEHSKAVLARSKLESICRELQRNRSIK 256
QY 241 EGVQARBEERKREKVTSHFQVTLNDIOLQMEQHNERNKSLRQNMELAEELKLIROY 300
DB 257 EGVQARBEERKREKVTSHFQVTLNDIOLQMEQHNERNKSLRQNMELAEELKLIROY 316
QY 301 ELREEHIDKVFHKDLQKQSQVQKDLHGEHSKAVLARSKLESICRELQRNRSIK 360
DB 317 ELREEHIDKVFHKDLQKQSQVQKDLHGEHSKAVLARSKLESICRELQRNRSIK 376
QY 361 KQETHLKOALALYTEKPEEFQNTLSKSEVFTTTPQEMEKMTKKIKKLEKTTMYRSRW 420
DB 377 KQETHLKOALALYTEKPEEFQNTLSKSEVFTTTPQEMEKMTKKIKKLEKTTMYRSRW 436
QY 421 ESSNKALLEMAEKTVDKELSGLOVKIQRLEKLCALQTERNDLNKRVQDLSAGGQSL 480
DB 437 ESSNKALLEMAEKTVDKELSGLOVKIQRLEKLCALQTERNDLNKRVQDLSAGGQSL 496
QY 481 TDSGPRRPEGGAQAPSPRVTEAPCPVGPAPSTASGQTGPQPTTSARA 530
DB 497 TDSGPRRPEGGAQAPSPRVTEAPCPVGPAPSTASGQTGPQPTTSARA 546

RESULT 4

Q86T85 ID Q86T85 PRELIMINARY; PRT; 546 AA.
AC Q86T85;
DT 01-JUN-2003 (TREMELrel. 24, Created)
DT 01-JUN-2003 (TREMELrel. 24, Last sequence update)
DE Hypothetical protein DKZP451J0118.
GN DKZP451J0118.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC Ansoorge W., Krieger S., Mewes H.W., Weil B., Amid C., Fobo G., Han M.,
RA Osanger A., Wiemann S.,
RL Submitted (APR-2003) to the EMBL/GenBank/DBJ databases.
DR EMBL; AL832637; CAD89952.1; -
KW Hypothetical protein.
SQ SEQUENCE 546 AA; 61846 MW; 36718BAE3AA7B6C2 CRC64;

Query Match 99.1%; Score 2677; DB 4; Length 546;
Best Local Similarity 99.2%; Pred. No. 6.1e-120;
Matches 526; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

QY 1 KSSPGQPEAGPEGAQERPSQAAPAVEAGPGSSQAAPRKEGAQARTAGSALRDVSEELS 60
DB 17 KSSPGQPEAGPEGAQERPSQAAPAVEAGPGSSQAAPRKEGAQARTAGSALRDVSEELS 76
QY 61 RQLEDILSTYCVNNQGGPGEAGQEPAPAEPAEDAKSRITYVARNGEPPTPVVYGEKPS 120
DB 77 RQLEDILSTYCVNNQGGPGEAGQEPAPAEPAEDAKSRITYVARNGEPPTPVVYGEKPS 136

QY 121 KGDPTNTEIRQSDVGDHRRPQB--KKKAGLKGKITLLMOTLNTLSTPEKLAALCKY 180
 DB |||||
 QY 137 KGDPTNTEIRQSDVGDHRRPQB--KKKAGLKGKITLLMOTLNTLSTPEKLAALCKY 196
 DB |||||
 QY 181 AELLLEHNSOKWMLLQKQKQOLVQKDHRLRGEHSHKAVLARSKLESCLRELQHNRSK 240
 DB |||||
 QY 197 AELLLEHNSOKWMLLQKQKQOLVQKDHRLRGEHSHKAVLARSKLESCLRELQHNRSK 256
 DB |||||
 QY 241 BEGVORAREBEERKEVTSHPFTVNDIOLQMEQHNRNSKLRQENMELAEKLI 300
 DB |||||
 QY 257 BEGVORAREBEERKEVTSHPFTVNDIOLQMEQHNRNSKLRQENMELAEKLI 316
 DB |||||
 QY 301 ELREEHIDKVPKHKDLOQOLVDAKLAQAQOEMLKEAEERHOREKDFLLKEAVESQRMCELM 360
 DB |||||
 QY 317 ELREEHIDKVPKHKDLOQOLVDAKLAQAQOEMLKEAEERHOREKDFLLKEAVESQRMCELM 376
 DB |||||
 QY 361 KQETHLKOQALALYATEKPEFONLTKSSSEVPTTFKQEMKMTKKIKLEKETTYWRSRW 420
 DB |||||
 QY 377 KQETHLKOQALALYATEKPEFONLTKSSSEVPTTFKQEMKMTKKIKLEKETTYWRSRW 436
 DB |||||
 QY 421 ESSNKALLEMAEBKTVRDKELEGLQVKIQRLKLCALQTERNDLNKRVQDLSAGGQSL 480
 DB |||||
 QY 437 ESSNKALLEMAEBKTVRDKELEGLQVKIQRLKLCALQTERNDLNKRVQDLSAGGQSL 496
 DB |||||
 QY 481 TDSGPRRPRGCAQAPSPRVTEAPCYPCAPSTEASGQTGPQETSARA 530
 DB |||||
 QY 497 TDSGPRRPRGCAQAPSPRVTEAPCYPCAPSTEASGQTGPQETSARA 546
 DB |||||

RESULT 5

Q8BHN1
 ID Q8BHN1 PRELIMINARY; PRT; 524 AA.
 AC Q8BHN1
 DT 01-MAR-2003 (T-EMBLrel. 23, Created)
 DT 01-MAR-2003 (T-EMBLrel. 23, Last sequence update)
 DT 01-OCT-2003 (T-EMBLrel. 25, Last annotation update)
 DE Hypothetical protein.
 GN RBBP7.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 OX NCBI_TaxID=10090;
 RN [1]
 RC SEQUENCE FROM N.A.
 RP STRAIN=C57BL/6J; TISSUE=Cortex, Head, and Testis;
 RX MEDLINE=22354683; PubMed=12466851;
 RA The FANTOM Consortium.
 RA the RIKEN Genome Exploration Research Group Phase I & II Team;
 RT "Analysis of the mouse transcriptome based on functional annotation of
 RT 60,770 full-length cDNAs."
 RL Nature 420:563-573 (2002).
 DR EMBL; AK030100; BAC26785.1; --
 DR EMBL; AK031783; BAC27547.1; --
 DR EMBL; AK044130; BAC31791.1; --
 DR MGD; MGI:1194910; Rbbp7.
 DR GO; GO:0005634; C:nucleus; IDA.
 DR GO; GO:0016564; P:transcriptional repressor activity; IDA.
 DR GO; GO:0001222; P:negative regulation of transcription from P...; IDA.
 KW Hypothetical protein.
 SQ SEQUENCE 524 AA; 60308 MW; 022877763334ED7C CRC64;

Query Match 47.8%; Score 1291; DB 11; Length 524;
 Best Local Similarity 57.5%; Pred. No. 4.8e-54;
 Matches 273; Conservative 70; Mismatches 90; Indels 42; Gaps 9;
 QY 10 GEGADERSQAPAVEAGPGSSQAPKPR---GAQARTAGSGLRDVSELSRQLEDI 66
 DB |||||
 QY 11 GGGGTEESAGGGRGRRSP-----POKFEIGTWEEARICGLGVKADM--VCMNQANDI 63
 DB |||||
 QY 67 LSTCYVDNNOGG-----PGBDGAGGPAEPDEAKSRVVARNGEPEPTPVVYGEKEP 119
 DB |||||
 QY 64 LQHQ--DPSCGGTTKHSLEGGEGS-----DFITKNRLVSS-----VFCTQEK 105
 DB |||||

QY 120 SKGDPNTEIRQSDVGDHRRPQB--KKKAGLKGKITLLMOTLNTLSTPEKLAALC 177
 DB |||||
 QY 106 RESIPGREA-----RTGPPDGGQDSCSRNKEKTLGKEVLLMQALNTLSTPEKLAALC 160
 DB |||||
 QY 178 KCYAEILLLEHNSOKWMLLQKQKQOLVQKDHRLRGEHSHKAVLARSKLESCLRELQHNRS 237
 DB |||||
 QY 161 KYADILLEHNSVQKMLLQKQKQOLVQKDHRLRGEHSHKAVLARSKLESCLRELQHNRS 220
 DB |||||
 QY 238 SLKEGVORAREBEERKEVTSHPFTVNDIOLQMEQHNRNSKLRQENMELAEKLI 297
 DB |||||
 QY 221 TLKEENNQAAREBEERKEATAHFOITLNEIQALEQHDHNAKLQENIELGEKLCUI 280
 DB |||||
 QY 298 EQVELREEHIDKVPKHKDLOQOLVDAKLAQAQOEMLKEAEERHOREKDFLLKEAVESQRMCELM 357
 DB |||||
 QY 281 EQVALREEHIDKVPKHKDLOQOLVDAKLAQAQOEMLKEAEERHOREKDFLLKEAVESQRMCELM 340
 DB |||||
 QY 358 ELKQKQETHLKOQALALYATEKPEFONLTKSSSEVPTTFKQEMKMTKKIKLEKETTYWRSRW 417
 DB |||||
 QY 341 EQKQKQETHLKOQALALYATEKPEFONLTKSSSEVPTTFKQEMKMTKKIKLEKETTYWRSRW 400
 DB |||||
 QY 418 SRNESNKALLEMAEBKTVRDKELEGLQVKIQRLKLCALQTERNDLNKRVQDLSAGGQSL 472
 DB |||||
 QY 401 TKWENNKALLEMAEBKTVRDKELEGLQVKIQRLKLCALQTERNDLNKRVQDLSAGGQSL 455
 DB |||||

RESULT 6

Q9NUQ3
 ID Q9NUQ3 PRELIMINARY; PRT; 528 AA.
 AC Q9NUQ3
 DT 01-OCT-2000 (T-EMBLrel. 15, Created)
 DT 01-OCT-2000 (T-EMBLrel. 15, Last sequence update)
 DT 01-OCT-2002 (T-EMBLrel. 22, Last annotation update)
 DE Hypothetical protein FLJ11209.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RC SEQUENCE FROM N.A.
 RP TISSUE=Placenta.
 RA Isogai T., Ota T., Hayaashi K., Sugiyama T., Otsuki T., Suzuki Y.,
 RA Nishikawa T., Nagai K., Sugano S., Shiratori A., Sudo H.,
 RA Kagatsuma M., Hosoi T., Kaku Y., Kodaira H., Kondo H., Sugawara M.,
 RA Takahashi M., Chiba Y., Ishida S., Murakawa K., Ono Y., Takiguchi S.,
 RA Matanabe S., Kimura K., Murakami K., Ishii S., Kawai Y., Saifu K.,
 RA Yamamoto J., Wakamatsu A., Nakamura Y., Nagahara K., Masuho Y.,
 RA Ninomiya K., Iwayanagi T.;
 RT "NEDO human cDNA sequencing project."
 RL Submitted (FEB-2000) to the EMBL/GenBank/DBSJ databases.
 DR EMBL; AK002071; BAA92068.1; --
 DR Genes; HGNC:18578; CXorf15.
 KW Hypothetical protein.
 SQ SEQUENCE 528 AA; 60605 MW; CAE8D781D806ACB3 CRC64;

Query Match 47.5%; Score 1283.5; DB 4; Length 528;
 Best Local Similarity 53.4%; Pred. No. 1.1e-53;
 Matches 285; Conservative 75; Mismatches 131; Indels 43; Gaps 10;
 QY 12 EGAQERPSQAPAVEAGPGSSQAPK--PGQARTAGSGLRDVSELSRQLEDI 69
 DB |||||
 QY 7 EAARGGGAEEATEAGRGRRSRPRKFEIGTWEEAGICGLGVKADMCLNSQSNILQH 66
 DB |||||
 QY 70 YCVDDNNOGGPGSDGAGGPAEPDEAKSRVVARNGEPEPTPVVYGEKEPSKGDPTTEEI 129
 DB |||||
 QY 67 Q--GSNGGTSNKHSLDEEGSDFTENNLVS-----PAYCTQSESEETPGG----EA 114
 DB |||||
 QY 130 ROSDVGDDHRRPQBKKKAGLKGKITLLMOTLNTLSTPEKLAALCKYAEILLLEHNS 189
 DB |||||
 QY 115 RTDPPDGGQDSCSRNKEKTLGKEVLLMQALNTLSTPEKLAALCKYAEILLLEHNS 172
 DB |||||
 QY 190 SOKWMLLQKQKQOLVQKDHRLRGEHSHKAVLARSKLESCLRELQHNRSKLEEGVQARE 249
 DB |||||
 QY 173 VQKQKMLLQKQKQOLVQKDHRLRGEHSHKAVLARSKLESCLRELQHNRSKLEEGVQARE 232
 DB |||||

```
QY 250 EEEKREKVTSHFQVTLNDIOLQMEQHNRNSKLRQENMLAERLKLBOYHLREHIDK 309
Db 233 EBERREKETAHFQITLDEIQAOLQEHDIENAKLRQENIELGKELKLBOYALREHIDK 292
QY 310 VFQKDLQOOLVDKLOQAQOEMLKEAERHOREKDFLLXEAVESQRMCELMKQOETHLQ 369
Db 293 VFQKDLQOOLVDKLOQAQOEMLKEAERHOREKDFLLXEAVESQRMCELMKQOETHLQ 352
QY 370 QALYTEKFEFQONTLSKSEVPTTFKQEKERTKIKLEKETTMYRWRSSNKALLR 429
Db 353 QLSLYMDKFEFQONTLSKSEVPTTFKQEKERTKIKLEKETTMYRWRSSNKALLR 412
QY 430 MABEKTROKELGLOVKIORLEKLCRALQOTERNDLNKRVO-----DL 472
Db 413 MABEKTROKELGLOVKIORLEKLCRALQOTERNDLNKRVO-----DL 472
QY 473 SAGGQSLT--DSGPE-----RPEPGQAQ--PSSPR--VTEAPCPGAPSTE 516
Db 473 ATPVMPCTALDSKSLNTSSKALGAHLEAPKQSAVQKPPSTGSAPEIES 526
RESULT 7
Q86T52 ID Q86T52 PRELIMINARY; PRT; 718 AA.
AC Q86T52;
DT 01-JUN-2003 (TREMELrel. 24, Created)
DT 01-JUN-2003 (TREMELrel. 24, Last sequence update)
DT 01-JUN-2003 (TREMELrel. 24, Last annotation update)
DE Hypothetical protein DKFP451G083 (Fragment).
GN DKFP451G083.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Human skeletal muscle;
RA Wambutt R., Heubner D., Mewes H.W., Weil B., Amid C., Oeanger A.,
RA Fobo G., Han M., Wiemann S.;
RL Submitted (MAY-2003) to the EMBL/GenBank/DBJ databases.
DR EMBL; AL831988; CAD91140.1; -.
KW Hypothetical protein.
FT NON_TER
SQ SEQUENCE 718 AA; 80290 MW; 16C3BF22C19559A4 CRC64;
Query Match 46.9%; Score 1266.5; DB 4; Length 718;
Best Local Similarity 46.4%; Pred. No. 9.6e-53;
Matches 276; Conservative 90; Mismatches 134; Indels 95; Gaps 10;
QY 18 PSQAAPAVEAR-----GPG-SSQAP-----RKPEGAQART-----AQSGALR 53
Db 28 PSDTTPKKBANHSQLSAERQSTPPGDSSSLPSHNGLEKEDGQDSPTVPQPEKASVHP 87
QY 54 DVSELSRQLEDILSTYCVNNQGGPGEDGAQGEPAEPDAEKSRITYVARNGEPPTPVV 113
Db 88 DISELNARQLEDIINTY--GSAASTAGKESASARASEQENAESPDN---EDGDCERTTEE 142
QY 114 YGEKEPSKGDPTNTEELRQSDVGDORHRRPOBKKAAGLGEITLLMQTLNTLSTPEKL 173
Db 143 AGREPVASGEPPTVK----EPVSNKEQK--LEKKILGKLGKKAANLLMQLNKQTPPEKF 196
QY 174 AALCKKYAELLBEHRNSQOMKLLQKKQSQVLQVQKDHLEHSGHSAVLARSKLSLCRELQ 233
Db 197 DFLPKKYAELLDEHRTQKKLLQKKQVQIQKEKQOLQGEHSRAILARSKLSLCRELQ 256
QY 234 RHNRSLKEGVQVQARBEERKKEVTSHTFQVTLNDIOLQMEQHNRNSKLRQENMLAERL 293
Db 257 RHNTLKEELQARBEERKKEVTSHTFQVTLNDIOLQMEQHNRNSKLRQENMLAERL 316
QY 294 KKLIEQVELREHIDKVFQKDLQOOLVDKLOQAQOEMLKEAERHOREKDFLLXEAVES 353
Db 317 KSIIDQVELREHIDKVFQKDLQOOLVDKLOQAQOEMLKEAERHOREKDFLLXEAVES 376
```

```
QY 354 QRMCELMKQOETHLKOOLALYTEKFEFQONTLSKSEVPTTFKQEKERTKIKLEKET 413
Db 377 KLOAKVLEQOETHLQOALYTEKFEFQONTLSKSEVPTTFKQEKERTKIKLEKET 436
QY 414 TWYRSWRSSNKALLREKERTKIKLEKETTMYRWRSSNKALLREKERTKIKLEKET 473
Db 437 ATWKARFENCNKALIDMTERRKALAKKEYECFVMKIGRLNLCRALQOBERNLHKKIRDAE 496
QY 474 AGGQSLTDSGPERPEG----- 491
Db 497 ISEKDDQSHNSDEEPESNVSDQIEDAEVNSVQTAVKNLATAFMIIHHPSTPHQSKE 556
QY 492 -----PQAQAPSSP-----RVTEAPCPGAPSTEASQOTGPQPTSA 528
Db 557 TQPEIGSSQESADAALKEPEQPLIPSRDSSEPLPLTPQAEAGCGSDAEPESKA 611
RESULT 8
Q8N3L3 ID Q8N3L3 PRELIMINARY; PRT; 715 AA.
AC Q8N3L3;
DT 01-OCT-2002 (TREMELrel. 22, Created)
DT 01-JUN-2003 (TREMELrel. 24, Last sequence update)
DT 01-JUN-2003 (TREMELrel. 24, Last annotation update)
DE Hypothetical protein DKFP451F022 (Fragment).
GN DKFP451F022.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC Wambutt R., Heubner D., Mewes H.W., Weil B., Amid C., Fobo G., Han M.,
RA Oeanger A., Wiemann S.;
RL Submitted (APR-2003) to the EMBL/GenBank/DBJ databases.
DR EMBL; AL834248; CAD38924.2; -.
KW Hypothetical protein.
FT NON_TER
SQ SEQUENCE 715 AA; 79798 MW; 11C50191BCD26582 CRC64;
Query Match 46.8%; Score 1265.5; DB 4; Length 715;
Best Local Similarity 46.4%; Pred. No. 1.1e-52;
Matches 276; Conservative 90; Mismatches 134; Indels 95; Gaps 10;
QY 18 PSQAAPAVEAR-----GPG-SSQAP-----RKPEGAQART-----AQSGALR 53
Db 25 PSDTTPKKBANHSQLSAERQSTPPGDSSSLPSHNGLEKEDGQDSPTVPQPEKASVHP 84
QY 54 DVSELSRQLEDILSTYCVNNQGGPGEDGAQGEPAEPDAEKSRITYVARNGEPPTPVV 113
Db 85 DISELNARQLEDIINTY--GSAASTAGKESASARASEQENAESPDN---EDGDCERTTEE 139
QY 114 YGEKEPSKGDPTNTEELRQSDVGDORHRRPOBKKAAGLGEITLLMQTLNTLSTPEKL 173
Db 140 AGREPVASGEPPTVK----EPVSNKEQK--LEKKILGKLGKKAANLLMQLNKQTPPEKF 193
QY 174 AALCKKYAELLBEHRNSQOMKLLQKKQSQVLQVQKDHLEHSGHSAVLARSKLSLCRELQ 233
Db 194 DFLPKKYAELLDEHRTQKKLLQKKQVQIQKEKQOLQGEHSRAILARSKLSLCRELQ 253
QY 234 RHNRSLKEGVQVQARBEERKKEVTSHTFQVTLNDIOLQMEQHNRNSKLRQENMLAERL 293
Db 254 RHNTLKEELQARBEERKKEVTSHTFQVTLNDIOLQMEQHNRNSKLRQENMLAERL 313
QY 294 KKLIEQVELREHIDKVFQKDLQOOLVDKLOQAQOEMLKEAERHOREKDFLLXEAVES 353
Db 314 KSIIDQVELREHIDKVFQKDLQOOLVDKLOQAQOEMLKEAERHOREKDFLLXEAVES 373
QY 354 QRMCELMKQOETHLKOOLALYTEKFEFQONTLSKSEVPTTFKQEKERTKIKLEKET 413
Db 374 KLOAKVLEQOETHLQOALYTEKFEFQONTLSKSEVPTTFKQEKERTKIKLEKET 433
```

QY 414 TWYGRWSSNKALLEAEKTVTKDEGLGVNKKIQRLEKLCRALQTERNDLNGRVODLUS 477
 :::|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
 Db 434 ATWKARFENCNKALDMDTEEKALRAKYECFVMKIGRLCNLCRALQERNEHFKCIDAE 493
 QY 474 AGQGSLTDSGPERPEG----- 491

Db 494 ISEKDDQSCHNSDEPESNVSDQIDAEVNSVTAVKMLATAFMIIHPSTPHQSKE 553

QY 492 -----PGAQAFSP-----RVTAPCYPCAPSTEASGGTGPOEPTSA 528

Db 554 TQPEIGSSQESADAALKPEQPPLIPSRDSSPLPLTPQAEGSGDAEPPSKA 608

RESULT 9

ID	Q91969	PRELIMINARY;	PRT; 676 AA.
AC	Q91969;		
DT	01-OCT-2000 (TrEMBLrel. 15, Created)		
DT	01-OCT-2000 (TrEMBLrel. 15, Last sequence update)		
DT	01-JUN-2003 (TrEMBLrel. 24, Last annotation update)		
DE	Muscle derived protein.		
GN	MDP77.		
OS	Gallus gallus (Chicken).		
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
OC	Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;		
OC	Gallus.		
OX	NCBI_TaxID=9031;		
RN	[1]		
RP	SEQUENCE FROM N.A.		
RC	TISSUE=Muscle;		
RX	MEDLINE=20175243; PubMed=10708594;		
RA	Uveda A., Fukui I., Fujimori K., Kiyosue K., Nishimune H., Kasai M.,		
PA	Taguchi T.;		
RT	*MDP77: A novel neurite-outgrowth-promoting protein predominantly		
RT	expressed in chick muscles.*;		
RL	Biochem. Biophys. Res. Commun. 269:564-569 (2000).		
DR	EMBL; D89999; BAA94755.1; --		
DR	PIR; JCT222; JCT222.		
SQ	SEQUENCE 676 AA; 77020 MW; FCEA9E393250BE94 CRC64;		

Query Match 45.6%; Score 1233; DB 13; Length 676;
 Best Local Similarity 53.9%; Pred.No. 3.6e-51;
 Matches 255; Conservative 84; Mismatches 100; Indels 34; Gaps 7

QY 2 SSPQQPAGPEGGAQRPSSQAAVAFAEAGSGSQAPRKPEGAQTAAQSGALRDVSELSR 61

Db 15 TSPTQDNQ-----QSKAFVPVPSLP--TNQTSAQFMATC-----DISELNR 59

QY 62 QLIEDILSTYCVDNNGGPGEDGA---QGEPAEPDPAEKSYTYVARNGEPPTPVYGEK 117

Db 60 QLEDILTNY-----GSAASLVKEGTTATDPKEKHVGSMEDAECEDVNHESEKD 110

QY 118 EPSKGDPNTEBIRQSDVEVDGRDHRRPQEEKKAGLGKEITLLMOTNLTLSTPEKLAAIC 177

Db 111 KPAPGDAS----RANFPSASKGK--LEKKILKGLGKEATILLQSLNKLITPEKLDLLF 164

QY 178 KKYAELLEHRNSCKQMELLQKQSOVLQVKDHLRGHSKAVLARSKLESICRELQHNR 237

Db 165 KKVAELLEHRAEQKQLKYLKQEAQITKPKDLQASEHSRAILARSKLESICRELQHNK 224

QY 238 SLKKEGVORAREEEERRKKTWSHFQVTLNDIQLOHQEQNRNRSKLQRNMELAEELAKLI 297

Db 225 TLKEETTORAREBDEKRKBETINHFQGTLSIEIQALEEQSERNMKLCQNTTELAKLSII 284

QY 298 EQYELREEHDKVFHKHLDQOOVLADAKLQAOEMLKEAEERHQREFKDFLLKEAVESQMC 357

Db 285 DQYELREHLDKLFKHELQQLVDAKLQSQEMMKAEERHQEKEYLLNQAAEWKLQA 344

QY 358 ELMKQOETHLKKQALAHYTEKPEEFQNTLSKSSEVPTTFQENKQTKIKLKEKTTMYR 417

Db 345 KWLKEQETVLQAQITLYYSERFESFQTKLTAKSNEVFATTQENKQTKKKKKLEKDTATWK 404

QY 418 SPWFSSNKAILLEAEKTVTKDEGLGVNKKIQRLEKLCRALQTERNDLNGRVODLUS 470

RESULT 11	
Q8BUK2	PRELIMINARY; PRT; 463 AA.
AC	Q8BUK2;
DT	01-MAR-2003 (TrEMBLrel. 23, Created)
DT	01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT	01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE	Muscle-derived protein XBP77 variant 1 (Fragment).
GN	2310001N14RIK.
OS	Mus musculus (Mouse).
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX	NCBI_TaxID=10090;
RN	[1]
RP	SEQUENCE FROM N.A.
RC	STRAIN=C57BL/6J; TISSUE=Heart;
RX	MEDLINE=22354683; PubMed=12466851;
RA	The PANTOM Consortium,
RA	the RIKEN Genome Exploration Research Group Phase I & II Team;
RT	"Analysis of the mouse transcriptome based on functional annotation of
RT	60,770 full-length cDNAs."
RL	Nature 420:563-573(2002).
DR	EMBL; AK084639; BAC39238.1; -
DR	MGI; MGI:1916756; 2310001N14RIK.
FT	NON TER 463
SQ	SEQUENCE 463 AA; 54090 MW; 18E9F881D73D59AA CRC64;
Query Match 44.2%; Score 1194; DB 11; Length 463;	
Best Local Similarity 52.7%; Pred. No. 1.7e-49;	
Matches 251; Conservative 72; Mismatches 103; Indels 50; Gaps 6;	
QY	18 PSQAAPAVEAGSGSQAPKPGQAQRTAQSGAL-----RDVSELSRQLSDILSTYC 71
DB	16 PPGDSSSLNQPGKQDGCSTGQA-PPQEGSLHPEKGAHDVABELSRQLSDIITY- 73
QY	72 VDNNGGPGDGAQGPAPEDAEKSTYVARNGEPEPTPVVYGEKPSKDPNTEIRQ 131
DB	74 -----GSAAPRGKSTSETKQPPNTEAPDNEDVDYEE-----TTBEI-- 112
QY	132 SDEVGDRHRPQ-----EKKAKGLGKEITLLMQTLNTLSTPEKLA 175
DB	113 -----DREPTAPEAPAAKEPVSKQEKLEKILGLGKEANLLQNLNKLQAPBEKLP 167
QY	176 LCKYAELEBEHNSQKMLQKQSQOLVQEKDHLRGESKAVLARSKLSLCRELQKH 235
DB	168 LFKKYTELDEHRTQKLLKQQAQTORQKQDQSEHNRVAVLARSKLSLCRELQKH 227
QY	236 NRSIKKEGVORAREEKKREKVTSHFOVTLNDIQLQMEQHNRNSKLRQENMELAEELKX 295
DB	228 NKLKEETLQAREEKKREKVTSHFOVTLNDIQLQMEQHNRNSKLRQENMELAEELKX 287
QY	296 LIBQYELREHDKVFKHDLQOLVDKLOQAQEMLKEABERHOREKDFLLKEAVESQR 355
DB	288 IIDQYELREHDKVFKHDLQOLVDKLOQAQEMLKEABERHOREKDFLLKEAVESQR 347
QY	356 MCELAKQOETHLKQALALYKTEKEFEFNTLSKSEVFTTQKQEMKMTKKIKLEKETT 415
DB	348 QAKVLKQOETHLKQALALYKTEKEFEFNTLSKSEVFTTQKQEMKMTKKIKLEKETT 407
QY	416 YRSRWSSNKLALMAEAKTVRDLKELGLQVKTQRLKLCRALQTERNDLNKRVQD 471
DB	408 WKARFENCNALLDMIEKALRAKEVECFVAKIQRLNLCRALQTERNDLNKRVQD 463
RESULT 12	
Q8N3S2	PRELIMINARY; PRT; 505 AA.
AC	Q8N3S2;
DT	01-OCT-2002 (TrEMBLrel. 22, Created)
DT	01-OCT-2002 (TrEMBLrel. 22, Last sequence update)
DT	01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
DE	Hypothetical protein (Fragment).
GN	DKFZP451A175.
OS	Homo sapiens (Human).
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX	NCBI_TaxID=9606;
RN	[1]
RP	SEQUENCE FROM N.A.
RC	Koehler K., Beyer A., Mewes H.W., Weil B., Wiemann S.;
RL	Submitted (JUL-2002) to the EMBL/GenBank/DBJ databases.
DR	EMBL; AL832322; CAD38617.1; -
KW	Hypothetical protein.
FT	NON TER 505
SQ	SEQUENCE 505 AA; 56957 MW; 1AA33548A2F1DADE CRC64;
Query Match 37.3%; Score 1008; DB 4; Length 505;	
Best Local Similarity 51.0%; Pred. No. 1.3e-40;	
Matches 203; Conservative 64; Mismatches 71; Indels 60; Gaps 2;	
QY	191 QKMKLLQKQSQOLVQEKDHLRGESKAVLARSKLSLCRELQKHNRSLKEGVORAREE 250
DB	1 QKMKLLQKQSQOLVQEKDHLRGESKAVLARSKLSLCRELQKHNRSLKEGVORAREE 60
QY	251 EKKRKEVTSHFQVTLNDIQLQMEQHNRNSKLRQENMELAEELKXLIBQYELREHDKV 310
DB	61 EKKRKEVTSHFQVTLNDIQLQMEQHNRNSKLRQENMELAEELKXLIBQYELREHDKV 120
QY	311 FKHKDLQOLVDKLOQAQEMLKEABERHOREKDFLLKEAVESQRMCELMKQOETHLKQ 370
DB	121 FKHKDLQOLVDKLOQAQEMLKEABERHOREKDFLLKEAVESQRMCELMKQOETHLKQ 180
QY	371 LALYKTEKEFEFNTLSKSEVFTTQKQEMKMTKKIKLEKETTMYRSWSSNKLALMA 430
DB	181 LTLYSGRFEFQSTLTLSKSEVFTTQKQEMKMTKKIKLEKETTMYRSWSSNKLALMA 240
QY	431 ABSEKTVRDKLEGLQVKTQRLKLCRALQTERNDLNKRVQDLSAGQSLTDSGPERRPE 490
DB	241 IBEKALRAKEVECFVAKIQRLNLCRALQTERNDLNKRVQDLSAGQSLTDSGPERRPE 300
QY	491 G-----PGAQ 495
DB	301 SNVSVQDEIDAEVNSVQTAVKALATAPWIIHPSTHQSKETQPEIGSSQESADAALK 360
QY	496 APSSP-----RVTEAPCYPGAPSTRAAGTQGPQPTSA 528
DB	361 EPEQPLIPSRDSESPLPLTPQABAGSGSDAEPSSKA 398
RESULT 13	
Q8BP11	PRELIMINARY; PRT; 241 AA.
AC	Q8BP11;
DT	01-MAR-2003 (TrEMBLrel. 23, Created)
DT	01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT	01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE	Hypothetical protein.
GN	RBBP7.
OS	Mus musculus (Mouse).
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX	NCBI_TaxID=10090;
RN	[1]
RP	SEQUENCE FROM N.A.
RC	STRAIN=C57BL/6J; TISSUE=Mullerian duct includes surrounding region;
RX	MEDLINE=22354683; PubMed=12466851;
RA	The PANTOM Consortium,
RA	the RIKEN Genome Exploration Research Group Phase I & II Team;
RT	"Analysis of the mouse transcriptome based on functional annotation of
RT	60,770 full-length cDNAs."
RL	Nature 420:563-573(2002).
DR	EMBL; AK078477; BAC37296.1; -
DR	MGI; MGI:1194910; Rbbp7.
DR	GO; GO:0005634; C:nucleus; IDA.
DR	GO; GO:0016564; P:transcriptional repressor activity; IDA.
DR	GO; GO:0000122; P:negative regulation of transcription from P...; IDA.

Db 106 KNESSLQIKVEERRKESQTKFOSSLNDVQKSLAKNNEIKLDRYNIEMTKLKLARQ 165
Qy 300 YELRESHIDKVPKHDLQOOLVDAKLQOQAEMLKEAERHOREKDFLLKEAVESQRMCEL 359
Db 166 YQTRSOHLEKLNQVQLEAQLQOQAEMLKEAERHOREKDFLLKEAVESQRMCEL 225
Qy 360 MKQOETHLQOALYTKSEFPNTLSKSSEVFTTFKQENBOMTKKIKLEKFTMYRSR 419
Db 226 LTRDEHQLKXQLNYTAKYDDFQOQLQKSNVFGSKVELEKMSKTKKIEREALQWRQK 285
Qy 420 WESSNKLALMAEKEKTVRDEHQLQKSNVFGSKVELEKMSKTKKIEREALQWRQK 471
Db 286 YEKANAMVIDLATKXSLQOQAEMLKEAERHOREKDFLLKEAVESQRMCEL 337

RESULT 16
Q86LQ1 PRELIMINARY; PRT; 264 AA.
AC Q86LQ1 (TREMELrel. 24, Created)
DT 01-JUN-2003 (TREMELrel. 24, Last sequence update)
DT 01-JUN-2003 (TREMELrel. 24, Last annotation update)
DE Hypothetical muscle-derived protein.
OS Branchiostoma belcheri (Amphioxus).
OC Eukaryota; Metazoa; Chordata; Cephalochordata; Branchiostomidae;
OC Branchiostoma.
OX NCBI_TaxID=7741;
RN [1]
R2 SEQUENCE FROM N.A.
RA Zhang H., Lin Y.
RL Submitted (DEC-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; AY191780; AA045170.1; -.
KW Hypothetical protein.
SQ SEQUENCE 264 AA; 31199 MW; EA0B82C96708815B CRC64;

Query Match 18.6%; Score 503; DB 5; Length 264;
Best Local Similarity 51.2%; Pred. No. 7.2e-17;
Matches 103; Conservative 33; Mismatches 57; Indels 8; Gaps 1;
Qy 282 LRQNMELAEKLIQOYELREHIDKVPKHDLQOOLVDAKLQOQAEMLKEAERHOREKDFLLKEAVESQRMCEL 341
Db 1 MKEENMELCGKLSLVEQYERKEHEVEKLINQKDLERQLAELQOQAEMLKEAERHOREKDFLLKEAVESQRMCEL 60
Qy 342 EKDFLLKEAVESQRMCELQOETHLQOQAEMLKEAERHOREKDFLLKEAVESQRMCEL 401
Db 61 EKQVLLKENYHRSRQMAIQEENMELCGKLSLVEQYERKEHEVEKLINQKDLERQLAELQOQAEMLKEAERHOREKDFLLKEAVESQRMCEL 120
Qy 402 MTKKIKLEKFTMYRSRQMAIQEENMELCGKLSLVEQYERKEHEVEKLINQKDLERQLAELQOQAEMLKEAERHOREKDFLLKEAVESQRMCEL 461
Db 121 MTKKIKLEKFTMYRSRQMAIQEENMELCGKLSLVEQYERKEHEVEKLINQKDLERQLAELQOQAEMLKEAERHOREKDFLLKEAVESQRMCEL 180
Qy 462 RNDINKRVQDLSAGQSLTD 482
Db 181 RSD-----TEGGADKVT 193

RESULT 17
Q84VE3 PRELIMINARY; PRT; 436 AA.
AC Q84VE3 (TREMELrel. 24, Created)
DT 01-JUN-2003 (TREMELrel. 24, Last sequence update)
DT 01-JUN-2003 (TREMELrel. 24, Last annotation update)
DE Muscle derived-like protein.
OS Oryza sativa (japonica cultivar-group).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
OC Eubacteriobacteriota; Bacteriota; Proteobacteria; Gammaproteobacteria; Rhodospirillales; Rhodospirillum rubrum.
OX NCBI_TaxID=39947;
RN [1]
R2 SEQUENCE FROM N.A.
RA Cooper B., Hutchison D., Park S., Guimil S., Lugimbuhl P., Ellero C.,

RA Goff S., Glazebrook J.;
RT "Identification of rice (Oryza sativa) Proteins Linked to the Cyclin-
Mediated Regulation of the Cell Cycle";
RL Submitted (JAN-2003) to the EMBL/GenBank/DBJ databases.
DR EMBL; AY224541; AA072661.1; -.
SQ SEQUENCE 436 AA; 48148 MW; 91B0958F888E2328 CRC64;

Query Match 17.1%; Score 462.5; DB 10; Length 436;
Best Local Similarity 28.9%; Pred. No. 1e-14;
Matches 140; Conservative 83; Mismatches 156; Indels 105; Gaps 15;
Qy 12 EGAQRPSCAPAVEA-----EGSGSQAPKPEGAQARTAQSGALRDVSELSQLEBD 65
Db 2 EGS---PATRLPEADSLDPGFVSSAADQAPPPP-----ADAGPAPRS----- 42
Qy 66 ILSTYCVNNQGGPGEDGAQGEPA-----EPEDAERSRTYVARGPPE----- 108
Db 43 -LGLQDADATVGGGGDETLGAPPSTLASVAQDTLDAYSADALQSLTVGSSAABPERAL 101
Qy 109 -----PTPVYGEKEPSPKDPNTEIRQSDVGDHRRPQEKCKAKGLKGTITLLM 160
Db 102 GEPVADAGAPVPVADAKESK---ESSVVEQVESMAD-----QKVVIAGSG--- 145
Qy 161 QTLNLTSTPEEKLAALCKKYABLLHEHRNSQOMKLLQKKQ-SQLVQEKDHLRGEHSAV 219
Db 146 -----EQKRVKVKSVKEDRELFEALQAYHKVVAERDAAI 181
Qy 220 LARSKLESICRELQHNRLKEEGVQVQRAAREBEKEKEVTISHFQVTLNDIQLQK-EQHN 278
Db 182 AVKEKLESICREBQFQNKMLKEECRRVSTEGQMMELSDKFNNAIKDVSXLDQKNEC 241
Qy 279 NSKLRQNMELAEKLIQOYELREHIDKVPKHDLQOOLVDAKLQOQAEMLKEAERHOREKDFLLKEAVESQRMCEL 338
Db 242 IAQLENNI-LASKLADADQNIQOYELREHIDKVPKHDLQOOLVDAKLQOQAEMLKEAERHOREKDFLLKEAVESQRMCEL 298
Qy 339 HQREKDFLLKEAVESQRMCELQOETHLQOQAEMLKEAERHOREKDFLLKEAVESQRMCEL 398
Db 299 TQWQ-----LYADQVSLM-STEKNLRLQLAADGERFOQFQDALTSTNEVFETKKE 349
Qy 399 MEKWKIKIKLEKFTMYRSRQMAIQEENMELCGKLSLVEQYERKEHEVEKLINQKDLERQLAELQOQAEMLKEAERHOREKDFLLKEAVESQRMCEL 458
Db 350 MEKWKIKIKLEKFTMYRSRQMAIQEENMELCGKLSLVEQYERKEHEVEKLINQKDLERQLAELQOQAEMLKEAERHOREKDFLLKEAVESQRMCEL 409
Qy 459 QTER 462
Db 410 QABR 413

RESULT 18
Q8BWJ7 PRELIMINARY; PRT; 241 AA.
AC Q8BWJ7 (TREMELrel. 23, Created)
DT 01-MAR-2003 (TREMELrel. 23, Last sequence update)
DT 01-JUN-2003 (TREMELrel. 24, Last annotation update)
DE Muscle-derived protein MDP77 variant 1 (Fragment).
GN 2310001N14RIK.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
R2 SEQUENCE FROM N.A.
RA STRAIN=C57BL/6J; TISSUE=Heart;
RX MEDLINE=22354683; PubMed=12466851;
RA The PANTOM Consortium,
RA the RIKEN Genome Exploration Research Group Phase I & II Team;
RT "Analysis of the mouse transcriptome based on functional annotation of
60,770 full-length cDNAs";
RL Nature 420:563-573 (2002).
DR EMBL; AK052324; BAC34937.1; -.
DR MGI; MGI:1916756; 2310001N14Rik.
FT NON_TER 241 241


```
SQ SEQUENCE 241 AA; 27232 MW; 0B2DD1BD215630BB CRC64;
Query Match 17.0%; Score 459; DB 11; Length 241;
Best Local Similarity 44.1%; Pred. No. 8.1e-15;
Matches 112; Conservative 25; Mismatches 67; Indels 50; Gaps 6;

Qy 18 PSQAAPAVEAGPGSSQAPRPEGAQARTAQSGAL-----RVSELSRQLEDILISTYC 71
Db 16 PPGSSSLNONGPGKQDGERCSTGQA-PEQEGSLHPEKGAHDVABELSRQLEDIISTY- 73
Qy 72 VDNQGGPGGDAQGEAPAEPAEDAESKRTYVARNGEPEPTPVYGEKEPSKDPNTEIRQ 131
Db 74 -----GSAASPRGKESTSETEQPPNTPAPNEDVDYER-----TTBEI-- 112
Qy 132 SDEVGDHRRPQ-----EKKAAGLGKEITLLMOTLNTLSTPPEKLA 175
Db 113 -----DREPTAPEPAAAKEPVSNKBOLEKKILGLGKEANLLMQLNKLQAPPEKLD 167
Qy 176 LCKKYAELLERHSNOKMLQKQSLQVQKDLHGEHSKAVLARSKLESLESLCRELQRH 235
Db 168 LFKYITELDHRTQKLLKLLKQAGTQKQKQLOQSEHNRVAVLARSKLESLESLCRELQRH 227
Qy 236 NRSLEKEGVQARE 249
Db 228 NKTLEETLQARE 241

RESULT 19
Q9RDX7 PRELIMINARY; PRT; 404 AA.
AC Q9RDX7
DT 01-JUN-2002 (TrEMBLrel. 21, Created)
DT 01-JUN-2002 (TrEMBLrel. 21, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Hypothetical protein (At5G50840).
GN Arabidopsis thaliana (Mouse-ear cress).
OS Arabidopsis
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids;
OC eurosida II; Brassicales; Brassicaceae; Arabidopsi.
OX NCBI_TaxID=3702;
RN [1]
RP SEQUENCE FROM N.A.
RA Nguyen M., Karlin-Neumann G., Southwick A., Lam B., Miranda M.,
RA Palm C.J., Bowser L., Jones T., Banh J., Carninci P., Chen H.,
RA Cheuk R., Chung M.K., Hayaishiaki Y., Ishida J., Kamiya A., Kawai J.,
RA Kim C., Lin J., Liu S.X., Narusaka M., Pham P.K., Sakano H.,
RA Sakurai T., Satou M., Seki M., Shim P., Yamada K., Shinozaki K.,
RA Ecker J., Theologis A., Davis R.W.;
RA Submitted (FEB-2002) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RA Cheuk R., Chen H., Kim C.J., Shinn P., Bowser L., Carninci P.,
RA Dale J.M., Hayashizaki Y., Ishida J., Jones T., Kamiya A.,
RA Karlin-Neumann G., Kawai J., Lam B., Lin J., Miranda M., Narusaka M.,
RA Nguyen M., Onodera C.S., Palm C.J., Quach H.L., Sakurai T., Satou M.,
RA Seki M., Southwick A., Tortum M., Wong C., Wu H.C., Yamada K., Yu G.,
RA Shinozaki K., Davis R.W., Theologis A., Ecker J.R.;
RA "Arabidopsis ORF clones";
RT Submitted (JUN-2003) to the EMBL/GenBank/DBJ databases.
DR EMBL; AY081318; AAL91207.1;
DR EMBL; F1003857; AAP68296.1;
KW Hypothetical protein.
SQ SEQUENCE 404 AA; 46796 MW; 3C31C38A7E437DF6 CRC64;

Query Match 16.7%; Score 450; DB 10; Length 404;
Best Local Similarity 31.4%; Pred. No. 3.7e-14;
Matches 127; Conservative 89; Mismatches 152; Indels 36; Gaps 11;

Qy 82 DGAQGEPA-EPDASKSRFY-----VARNGEPEPTPVYGEKEPSKDPNTEIROSDVG 136
Db 21 DGAETPPLSPKTOBETTHHETVALEKTEKPTFFVPLCETDGNEDDEVADLIQBSKL- 79

SQ SEQUENCE 241 AA; 27232 MW; 0B2DD1BD215630BB CRC64;
Query Match 17.0%; Score 459; DB 11; Length 241;
Best Local Similarity 44.1%; Pred. No. 8.1e-15;
Matches 112; Conservative 25; Mismatches 67; Indels 50; Gaps 6;

Qy 18 PSQAAPAVEAGPGSSQAPRPEGAQARTAQSGAL-----RVSELSRQLEDILISTYC 71
Db 16 PPGSSSLNONGPGKQDGERCSTGQA-PEQEGSLHPEKGAHDVABELSRQLEDIISTY- 73
Qy 72 VDNQGGPGGDAQGEAPAEPAEDAESKRTYVARNGEPEPTPVYGEKEPSKDPNTEIRQ 131
Db 74 -----GSAASPRGKESTSETEQPPNTPAPNEDVDYER-----TTBEI-- 112
Qy 132 SDEVGDHRRPQ-----EKKAAGLGKEITLLMOTLNTLSTPPEKLA 175
Db 113 -----DREPTAPEPAAAKEPVSNKBOLEKKILGLGKEANLLMQLNKLQAPPEKLD 167
Qy 176 LCKKYAELLERHSNOKMLQKQSLQVQKDLHGEHSKAVLARSKLESLESLCRELQRH 235
Db 168 LFKYITELDHRTQKLLKLLKQAGTQKQKQLOQSEHNRVAVLARSKLESLESLCRELQRH 227
Qy 236 NRSLEKEGVQARE 249
Db 228 NKTLEETLQARE 241

RESULT 19
Q9RDX7 PRELIMINARY; PRT; 404 AA.
AC Q9RDX7
DT 01-JUN-2002 (TrEMBLrel. 21, Created)
DT 01-JUN-2002 (TrEMBLrel. 21, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Hypothetical protein (At5G50840).
GN Arabidopsis thaliana (Mouse-ear cress).
OS Arabidopsis
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids;
OC eurosida II; Brassicales; Brassicaceae; Arabidopsi.
OX NCBI_TaxID=3702;
RN [1]
RP SEQUENCE FROM N.A.
RA Nguyen M., Karlin-Neumann G., Southwick A., Lam B., Miranda M.,
RA Palm C.J., Bowser L., Jones T., Banh J., Carninci P., Chen H.,
RA Cheuk R., Chung M.K., Hayaishiaki Y., Ishida J., Kamiya A., Kawai J.,
RA Kim C., Lin J., Liu S.X., Narusaka M., Pham P.K., Sakano H.,
RA Sakurai T., Satou M., Seki M., Shim P., Yamada K., Shinozaki K.,
RA Ecker J., Theologis A., Davis R.W.;
RA Submitted (FEB-2002) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RA Cheuk R., Chen H., Kim C.J., Shinn P., Bowser L., Carninci P.,
RA Dale J.M., Hayashizaki Y., Ishida J., Jones T., Kamiya A.,
RA Karlin-Neumann G., Kawai J., Lam B., Lin J., Miranda M., Narusaka M.,
RA Nguyen M., Onodera C.S., Palm C.J., Quach H.L., Sakurai T., Satou M.,
RA Seki M., Southwick A., Tortum M., Wong C., Wu H.C., Yamada K., Yu G.,
RA Shinozaki K., Davis R.W., Theologis A., Ecker J.R.;
RA "Arabidopsis ORF clones";
RT Submitted (JUN-2003) to the EMBL/GenBank/DBJ databases.
DR EMBL; AY081318; AAL91207.1;
DR EMBL; F1003857; AAP68296.1;
KW Hypothetical protein.
SQ SEQUENCE 404 AA; 46796 MW; 3C31C38A7E437DF6 CRC64;

Query Match 16.7%; Score 450; DB 10; Length 404;
Best Local Similarity 31.4%; Pred. No. 3.7e-14;
Matches 127; Conservative 89; Mismatches 152; Indels 36; Gaps 11;

Qy 82 DGAQGEPA-EPDASKSRFY-----VARNGEPEPTPVYGEKEPSKDPNTEIROSDVG 136
Db 21 DGAETPPLSPKTOBETTHHETVALEKTEKPTFFVPLCETDGNEDDEVADLIQBSKL- 79
```


152 IQKEI--TLLMOTINTLTPB-EKLAALCKYAEILLEBHRNSOKOMKLLQKKSQVQSK 208
1 MGNFDEAALLKSLF--GVPAEKVKLIKLAEE--SEKQNAELKIKVLVDYDKVVKVXDL 56
209 DHLRGHSHKAVL-----ARSLKESLCRELQHRNLSKEEGVQVAREEERKEKVTSHFQV 263
57 TEKKLERNQVILLRTTEBAKSLKLELCRGLQKQANHQTRREACAKMKKLEVERGLAVEQLKV 116
264 TLNDIOLQMEHNERNSKLRQENMELAKLILQYE-----LREHIDKVKP 312
117 TLKDIEKTMABGRSKSDSLAEDNKKLSKFSIGHYEEKMKVVDQIQKKEKYBEEYK 176
313 HXDLOQLVDKIQQAQOEMLEKAEERHOREKDFLLKEAVESQRMCELMKQOETHLKQOLA 372
177 TNDLHILKILAKLESASQVKS-----GMEKDELAKIMLEETARVGGALKTEKALREQVQ 232
373 LYTEKPEPQNTLSKSSVFTTPFKOEMERMTKKIKLEKBTMTYMSRWSSNKALL----- 428
233 EYSAKYSELTSCLSKSNEAFDKDEISRVNCKQVKEKGLSYKCKSDEANKKVLVTM 292
429 ---EVAEEKTVDEKLEGLQVKIORLEKLCRALQ 459
293 TNOEVAEKATSDK-----KIQMLENLCLRALR 319
RESULT 23
Q9BZA4 PRELIMINARY; PRT; 181 AA.
AC Q9BZA4
DT 01-JUN-2001 (TREMELrel. 17, Created)
DT 01-JUN-2001 (TREMELrel. 17, Last sequence update)
DT 01-JUN-2003 (TREMELrel. 24, Last annotation update)
DE Lipopolysaccharide-specific response 5-like protein (Hypothetical
protein DKP2P451G0616).
GN DKP2P451G0616
OS Homo sapiens (Human)
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RA Kuroda-Kawaguchi T., Skaletsky H., Minx P.J., Brown L.G., Rozen S.,
Wilson R.K., Waterston R.H., Page D.C.;
RT "The DNA sequence of the human Y chromosome";
RL Submitted (DEC-2000) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RA Ansorge W., Krieger S., Mewes H.W., Weil B., Amid C., Fobo G., Han M.,
Oaanger A., Wiemann S.;
RL Submitted (APR-2003) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF332225; AKK13477.1; -
DR EMBL; AL832583; CAD89940.1; -
DR Genew; HGNC:18577; CYorf15B.
KW Hypothetical protein.
SQ SEQUENCE 181 AA; 20858 MW; 4BA8B830933012A1 CRC64;
Query Match 13.7%; Score 369.5; DB 4; Length 181;
Best Local Similarity 44.8%; Pred. No. 1.1e-10;
Matches 78; Conservative 31; Mismatches 46; Indels 19; Gaps 3;
360 MKQOETHLKQALYATEKFEFQNTLSKSSVFTTPFKOEMERMTKKIKLEKBTMTYMSRW 419
1 MKQSEALKEQLFLYMDKFEFQNTVAKTNELFTAFKQETKLTKKIKLEKEMVITYTK 60
420 WESSKALLEMAEKEKTVRDEKLEGLQVKIORLEKLCRALQOTERNLKNRY----- 469
61 WENNNTLQMAEKEKTRDKNVYFOIKLERLKYALQIERNELSEKGLKGLKGVSVX 120
470 ---QDLISAGQGSQSLTDSGPER-----RPEGPQAQAPSPR-VTEAPCYPGAPT 514
121 VADVLDLAVPYTHSCADLSDSNMLNTSSKRAPGVHLRADPKGMVNEKVCYSKALST 174

01-MAR-2003 (TREMELrel. 23, Created)
01-MAR-2003 (TREMELrel. 23, Last sequence update)
01-MAR-2003 (TREMELrel. 23, Last annotation update)
DE Hypothetical protein (Fragment).
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=Olfactory brain;
RX MEDLINE=22354683; PubMed=12466851;
RA The FANTOM Consortium,
RA the RIKEN Genome Exploration Research Group Phase I & II Team;
RT "Analysis of the mouse transcriptome based on functional annotation of
60,770 full-length cDNAs";
RL Nature 420:563-573(2002).
DR EMBL; AK078192; BAC37168.1; -
KW Hypothetical protein.
FT NON TER 150
SQ SEQUENCE 150 AA; 16490 MW; C444717503B42E49 CRC64;
Query Match 14.5%; Score 392.5; DB 11; Length 150;
Best Local Similarity 57.1%; Pred. No. 7.3e-12;
Matches 88; Conservative 9; Mismatches 36; Indels 21; Gaps 4;
2 SSSQPPAGPEGAQERSQAAPAVEAREGGSSQAPRKPEGAQARTAGSGLRDV----- 55
5 NTPERRAARGNGKQDRE-----PMDDPQQLLGRKPKVPOARLL--GRPGLELKLQLS 56
56 -----SSELROLEDILSYVDNNGGPGEDGAQGEPAEPDAEKSRITYVARNGEPEP 109
57 LGPSVTHLSRQLLEDILSYVDNNGGPAEAGAQGEPTPEDEKTSRITYVARNGEPEP 116
110 -TPVYGEKSPKGDPTNTEIRQSDVGDHRHR 142
117 GIPVNGEKETSGEPGTETIRASDEVGDRHR 150
RESULT 22
Q22666 PRELIMINARY; PRT; 335 AA.
AC Q22666
DT 01-NOV-1996 (TREMELrel. 01, Created)
DT 01-MAY-2000 (TREMELrel. 13, Last sequence update)
DT 01-OCT-2003 (TREMELrel. 25, Last annotation update)
DE T22C1.6 protein.
GN T22C1.6
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC Rhabditidae; Pelodidae; Caenorhabditis.
OX NCBI_TaxID=6239;
RN [1]
RP SEQUENCE FROM N.A.
RA McMuray A.A.;
RL Submitted (JUN-1996) to the EMBL/GenBank/DBJ databases.
RN [2]
RX MEDLINE=99069613; PubMed=9851916;
RA none;
RT "Genome sequence of the nematode C. elegans: A platform for
investigating biology";
RL Science 282:2012-2018(1998).
DR EMBL; Z75550; CAA99923.2; -
DR PIR; F87844; F87844.
DR PIR; T25110; T25110.
DR WormPep; T22C1.6; C2524002.
SQ SEQUENCE 335 AA; 38338 MW; DFC45011AD2FD2F5 CRC64;
Query Match 13.8%; Score 372.5; DB 5; Length 335;
Best Local Similarity 30.8%; Pred. No. 1.5e-10;
Matches 103; Conservative 70; Mismatches 120; Indels 41; Gaps 9;

```
RESULT 24
Q9FGD8 PRELIMINARY; PRT; 372 AA.
AC Q9FGD8;
DT 01-MAR-2001 (TrEMBLrel. 16, Created)
DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT 01-MAR-2001 (TrEMBLrel. 16, Last annotation update)
DE Genomic DNA, chromosome 5, TAC clone:K16E14.
OS Arabidopsis thaliana (mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids;
OC eurosids II; Brassicales; Brassicaceae; Arabidopsids.
OX NCBI_TaxID=3702;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Columbia;
RA Kameko T., Katoh T., Asamizu E., Sato S., Nakamura Y., Kotani H.,
RA Tabata S.;
RT "Structural analysis of Arabidopsis thaliana chromosome 5. XI.";
RL Submitted (APR-1999) to the EMBL/GenBank/DBJ databases.
DR EMBL; AB026637; BAB10777.1;
DR EMBL; AB025617; BAB10777.1; JOINED.
SQ SEQUENCE 372 AA; 43086 MW; A706PFC3C75576 CRC64;

Query Match 12.3%; Score 332; DB 10; Length 372;
Best Local Similarity 26.8%; Pred. No. 1.4e-08;
Matches 106; Conservative 81; Mismatches 136; Indels 72; Gaps 13;

QY 109 PTPVYGEKPSKGDENT-BEIRQSDVGDHRRPOEKKKAGLKGITLLMTLTL 167
D5 16 PDGFDGATEPLNSPKTQETTHETVAIEKTEPR-----TFP 55
QY 168 TPSEKLAALCK-----KYAELEEHNSQKQWKLQKQSLVQEKDHURG-----B 214
D5 56 VP-----LCETDGNEDDEADLQESIKLEFEQKEKSPISQYDVSIDLCADDIE 109
QY 215 HSKAVLRSKLESICRELQRN-----RSLKRGVQRAREEERKEVTSHP 261
D5 110 SRRLVRSVRVQRICSN-QRKSFWNLCSISKSLKELEBECKRVSTEGTILRSGLSTKP 168
QY 262 QVTLLNDIQLOM-EQHNERNKLRQEMELAEKLKLEIQYELREHIDKVFKHDLQOQL 320
D5 169 QDAIMDSIKLDEQKNESLTQLK-ENEMLRTKLHLADQFMLSQEQHEQLRKQKTLQI 227
QY 321 VDALKQEQEMLKEAEERHOREKDFLKEAVESQRCMLKQOETHLQOQLALYTEKPEE 380
D5 228 SALKIKOHEKJL--THEOSQMK-----VTADQVSQLL-STENKRLQLTSDGKFPQ 276
QY 381 FQNTLSKSEVPTFKQEMKMKKKIKLEKETTMYRSWRSSNKALLEMAEETVRDKE 440
D5 277 FQDALVKSNEVETFKQEDDKMSKALKELRKENAFKNTKSKSDITLIELVEERBLKL 336
QY 441 IEGQLVKIQRLEKLCALQTER-----NDLNKRVQ 470
D5 337 LEKTKQKQKLSLCSLSQAEKQKQKTNSTDSAVQ 371

RESULT 25
Q9VTUO PRELIMINARY; PRT; 1168 AA.
AC Q9VTUO;
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-OCT-2002 (TrEMBLrel. 22, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE CG32652 protein.
DE CG32652 OR CG1905.
OS Drosophila melanogaster (fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A.
```

```
RC STRAIN=Berkeley;
RX MEDLINE=20196006; PubMed=10731132;
RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA Sutton G.G., Mortman J.R., Randell M.D., Zhang Q., Chen L.X.,
RA Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,
RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.B.G.,
RA Abril J.F., Agbayani A., An H.-J., Andrews-Frankoch C., Baldwin D.,
RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA Beeson K.Y., Benos P.V., Beriman B.P., Bhandari D., Bolshakov S.,
RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brottier P.,
RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,
RA Foslcr C., Gabrielian A.E., Gaig N.S., Gelbart W.M., Glasser K.,
RA Gloddek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA Harris N.L., Harvey D., Helman T.J., Hernandez J.R., Houck J.,
RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA Markulov G., Milshina N.V., Moharry C., Morris J., Moshrefi A.,
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.W.,
RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA Reinart K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA Swirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA Wang Z.-Y., Wasarman D.A., Weinstein G.M., Weissenbach J.,
RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S.-Y., Yao Q.A.,
RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhu Q., Zheng L.,
RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhou S., Zhou X., Smith H.O.,
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT "The genome sequence of Drosophila melanogaster.";
RL Science 287:2185-2195(2000).

SEQUENCE FROM N.A.
Celniker S.E., Adams M.D., Krommiller B., Wan K.H., Holt R.A.,
Evans C.A., Gocayne J.D., Amanatides P.G., Brandon R.C., Rogers Y.,
Banzon J., An H., Baldwin D., Banzon J., Beeson K.Y., Busam D.A.,
Carlson J.W., Center A., Champe M., Davenport L.B., Dietz S.M.,
Dodson K., Dorsett V., Doup L.E., Doyle C., Dresnek D., Farfan D.,
Ferreira S., Frise E., Galle R.E., Garg N.S., George R.A.,
Gonzalez M., Houck J., Hoskins R.A., Hostin D., Howland T.J.,
Ibegwam C., Jalali M., Kruse D., Li P., Mattei B., Moshrefi A.,
McIntosh T.C., Moy M., Murphy B., Nelson C., Nelson K.A., Nunoo J.,
Pacleb J., Paragas V., Park S., Patel S., Pfeiffer B.,
Phouanavong S., Pittman G.S., Puri V., Richards S., Scheeler F.,
Stapleton M., Strong R., Swirskas R., Tector C., Tyler D.,
Williams S.M., Zaveri J.S., Smith H.O., Venter J.C., Rubin G.M.;
RT "Sequencing of Drosophila melanogaster genome.";
RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.

SEQUENCE FROM N.A.
Miera S., Crosby M.A., Matthews B.B., Bayraktaroglu L., Campbell K.,
Hradecky P., Huang Y., Kaminker J.S., Prochnik S.E., Smith C.D.,
Tupy J.L., Bergman C., Berman B., Carlson J.W., Celniker S.E.,
Clamp M., Drysdale R., Emmert D., Frise E., de Grey A., Harris N.,
Kronmiller B., Marshall B., Millburn G., Richter J., Russo S.,
Searle S.M.J., Smith E., Shu S., Smutniak F., Whitfield E.,
Ashburner M., Gelbart W.M., Rubin G.M., Mungall C.J., Lewis S.B.;
RT "Annotation of Drosophila melanogaster genome.";
RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.

SEQUENCE FROM N.A.
Adams M.D., Celniker S.E., Gibbs R.A., Rubin G.M., Venter C.J.;
RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
[5]
```

```
QY 61 RTAQSGLRDVSEELSRQLEDILSTYCVNNOGGPGEAGQEPAPEDAEKSRITYVARN 120
Db 61 RTAQSGLRDVSEELSRQLEDILSTYCVNNOGGPGEAGQEPAPEDAEKSRITYVARN 120
QY 121 GEPEPTPVVNGEKPSPKSGDPNTEIEIRQSDVGDHRRPQEKKAAGLKEITLLMQTLN 180
Db 121 GEPEPTPVVNGEKPSPKSGDPNTEIEIRQSDVGDHRRPQEKKAAGLKEITLLMQTLN 180
QY 181 TLTSTPEKLAALCKKYAELLBEHNSQKMKLLQKQSQQLVQEKDHLRGHSHKAVLARSK 240
Db 181 TLTSTPEKLAALCKKYAELLBEHNSQKMKLLQKQSQQLVQEKDHLRGHSHKAVLARSK 240
QY 241 LESLCRELQHRNRLSKBEGVQARABEEERKEVTSHFQVTLNDIQLQMEQHNEHNSKLRQ 300
Db 241 LESLCRELQHRNRLSKBEGVQARABEEERKEVTSHFQVTLNDIQLQMEQHNEHNSKLRQ 300
QY 301 ENMELAEERLKKLIEQYELREBEHIDKVFHKKDLQQLVDKQLQQAQEMLEAEERHOREKD 360
Db 301 ENMELAEERLKKLIEQYELREBEHIDKVFHKKDLQQLVDKQLQQAQEMLEAEERHOREKD 360
QY 361 FLLKEAVESQRMCELMKQOETHLQKQALYTEKFEFQNTLSKSEVFTTFKQEMEKMTK 420
Db 361 FLLKEAVESQRMCELMKQOETHLQKQALYTEKFEFQNTLSKSEVFTTFKQEMEKMTK 420
QY 421 KIKKLEKETTYMRSRWESSNKALLEMAEKTVRDKLEGLQVKIQRLEKLCRALQTERND 480
Db 421 KIKKLEKETTYMRSRWESSNKALLEMAEKTVRDKLEGLQVKIQRLEKLCRALQTERND 480
QY 481 LNKRVQDLSAGQSGSLTDSGPERPPEGPAQAPSSPRVTEAPCYPGAPSTEASQGTGPQE 540
Db 481 LNKRVQDLSAGQSGSLTDSGPERPPEGPAQAPSSPRVTEAPCYPGAPSTEASQGTGPQE 540
QY 541 PTSARA 546
Db 541 PTSARA 546
```

RESULT 2

```
US-10-671-242-44
; Sequence 44, Application US/10671242
; Publication No. US2004004049A1
; GENERAL INFORMATION:
; APPLICANT: Lees, Robert S.
; APPLICANT: Law, Simon W.
; APPLICANT: Arjona, Anibal A.
; TITLE OF INVENTION: NOVEL LOW DENSITY LIPOPROTEIN BINDING
; TITLE OF INVENTION: PROTEINS AND THEIR USE IN DIAGNOSING AND TREATING
; TITLE OF INVENTION: ATHEROSCLEROSIS
; FILE REFERENCE: 10797-004001
; CURRENT APPLICATION NUMBER: US/10/671,242
; CURRENT FILING DATE: 2003-09-24
; PRIOR APPLICATION NUMBER: US/09/616,289
; PRIOR FILING DATE: 2000-07-14
; PRIOR APPLICATION NUMBER: US 09/517,849
; PRIOR FILING DATE: 2000-03-02
; PRIOR APPLICATION NUMBER: US 08/979,608
; PRIOR FILING DATE: 1997-11-26
; PRIOR APPLICATION NUMBER: US 60/031,930
; PRIOR FILING DATE: 1996-11-27
; PRIOR APPLICATION NUMBER: US 60/048,547
; PRIOR FILING DATE: 1997-06-03
; NUMBER OF SEQ ID NOS: 53
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 44
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-671-242-44
```

Query Match 100.0%; Score 546; DB 12; Length 546;
Best Local Similarity 100.0%; Pred. No. 0;

```
Matches 546; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 MNKQDKNGKAQKSNPKSPGQPEAGPEGAQERPSQAAPAVEAEGPGSSQAPRKPEGAQA 60
Db 1 MNKQDKNGKAQKSNPKSPGQPEAGPEGAQERPSQAAPAVEAEGPGSSQAPRKPEGAQA 60
QY 61 RTAQSGLRDVSEELSRQLEDILSTYCVNNOGGPGEAGQEPAPEDAEKSRITYVARN 120
Db 61 RTAQSGLRDVSEELSRQLEDILSTYCVNNOGGPGEAGQEPAPEDAEKSRITYVARN 120
QY 121 GEPEPTPVVNGEKPSPKSGDPNTEIEIRQSDVGDHRRPQEKKAAGLKEITLLMQTLN 180
Db 121 GEPEPTPVVNGEKPSPKSGDPNTEIEIRQSDVGDHRRPQEKKAAGLKEITLLMQTLN 180
QY 181 TLTSTPEKLAALCKKYAELLBEHNSQKMKLLQKQSQQLVQEKDHLRGHSHKAVLARSK 240
Db 181 TLTSTPEKLAALCKKYAELLBEHNSQKMKLLQKQSQQLVQEKDHLRGHSHKAVLARSK 240
QY 241 LESLCRELQHRNRLSKBEGVQARABEEERKEVTSHFQVTLNDIQLQMEQHNEHNSKLRQ 300
Db 241 LESLCRELQHRNRLSKBEGVQARABEEERKEVTSHFQVTLNDIQLQMEQHNEHNSKLRQ 300
QY 301 ENMELAEERLKKLIEQYELREBEHIDKVFHKKDLQQLVDKQLQQAQEMLEAEERHOREKD 360
Db 301 ENMELAEERLKKLIEQYELREBEHIDKVFHKKDLQQLVDKQLQQAQEMLEAEERHOREKD 360
QY 361 FLLKEAVESQRMCELMKQOETHLQKQALYTEKFEFQNTLSKSEVFTTFKQEMEKMTK 420
Db 361 FLLKEAVESQRMCELMKQOETHLQKQALYTEKFEFQNTLSKSEVFTTFKQEMEKMTK 420
QY 421 KIKKLEKETTYMRSRWESSNKALLEMAEKTVRDKLEGLQVKIQRLEKLCRALQTERND 480
Db 421 KIKKLEKETTYMRSRWESSNKALLEMAEKTVRDKLEGLQVKIQRLEKLCRALQTERND 480
QY 481 LNKRVQDLSAGQSGSLTDSGPERPPEGPAQAPSSPRVTEAPCYPGAPSTEASQGTGPQE 540
Db 481 LNKRVQDLSAGQSGSLTDSGPERPPEGPAQAPSSPRVTEAPCYPGAPSTEASQGTGPQE 540
QY 541 PTSARA 546
Db 541 PTSARA 546
```

RESULT 3

```
US-10-023-529-44
; Sequence 44, Application US/10023529
; Publication No. US20020129388A1
; GENERAL INFORMATION:
; APPLICANT: Lees, Robert S.
; APPLICANT: Law, Simon W.
; APPLICANT: Arjona, Anibal A.
; TITLE OF INVENTION: NOVEL LOW DENSITY LIPOPROTEIN BINDING
; TITLE OF INVENTION: PROTEINS AND THEIR USE IN DIAGNOSING AND TREATING
; TITLE OF INVENTION: ATHEROSCLEROSIS
; FILE REFERENCE: 10797-004001
; CURRENT APPLICATION NUMBER: US/10/023,529
; CURRENT FILING DATE: 2001-12-17
; PRIOR APPLICATION NUMBER: 09/616,289
; PRIOR FILING DATE: 2000-07-14
; PRIOR APPLICATION NUMBER: US 09/517,849
; PRIOR FILING DATE: 2000-03-02
; PRIOR APPLICATION NUMBER: US 08/979,608
; PRIOR FILING DATE: 1997-11-26
; PRIOR APPLICATION NUMBER: US 60/031,930
; PRIOR FILING DATE: 1996-11-27
; PRIOR APPLICATION NUMBER: US 60/048,547
; PRIOR FILING DATE: 1997-06-03
; NUMBER OF SEQ ID NOS: 53
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 44
; LENGTH: 546
; TYPE: PRT
```

; ORGANISM: Homo sapiens
US-10-023-529-44

Query Match 100.0%; Score 546; DB 13; Length 546;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 546; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

1 MKNQDKKNGAAGKSNPKSPGQPEAGPEGAQERPSQAAPAVEAGPGSSQAAPKPEGAQA 60
1 MKNQDKKNGAAGKSNPKSPGQPEAGPEGAQERPSQAAPAVEAGPGSSQAAPKPEGAQA 60

61 RTAQSGALRDVSEELSRQLEDILSTYCVDNNGQGGEDGAQGEPAPEPADEAKSRRTYVARN 120
61 RTAQSGALRDVSEELSRQLEDILSTYCVDNNGQGGEDGAQGEPAPEPADEAKSRRTYVARN 120

121 GEPEPTPVNGEKEPSKGDPTNTEIRQSDVEGDRDHRPQEKKAKGLGKEITLLMOTLN 180
121 GEPEPTPVNGEKEPSKGDPTNTEIRQSDVEGDRDHRPQEKKAKGLGKEITLLMOTLN 180

181 TLSTPEKLAALCKKYAELLBEHRNSQKQMLKKQKQSQLVQEKDHLRGEHSKAVLARSK 240
181 TLSTPEKLAALCKKYAELLBEHRNSQKQMLKKQKQSQLVQEKDHLRGEHSKAVLARSK 240

241 LESLCELQHNRSLSKEGVQVAREEKEKKEVTSHFQVTLNDIQLQMEQHNRNSKLQ 300
241 LESLCELQHNRSLSKEGVQVAREEKEKKEVTSHFQVTLNDIQLQMEQHNRNSKLQ 300

301 ENMELAEKLKLIQEVYELREEHIDKVFKHQDLQQLVDAKLQQAQEWLKEAEERHOREKD 360
301 ENMELAEKLKLIQEVYELREEHIDKVFKHQDLQQLVDAKLQQAQEWLKEAEERHOREKD 360

361 FLKKEAVESQRMCELMKQOETHLKOOLALYTEKPEFQNTLSKSSSEVFTTFKQEMEKMTK 420
361 FLKKEAVESQRMCELMKQOETHLKOOLALYTEKPEFQNTLSKSSSEVFTTFKQEMEKMTK 420

421 KIKKLEKTTMYRSRWESSNKALLEMAEKTVDKELQVGIQRLKLCRALQTERND 480
421 KIKKLEKTTMYRSRWESSNKALLEMAEKTVDKELQVGIQRLKLCRALQTERND 480

481 LNKRVQDLSAGGQSLTDSGPERPPEGPAQAPSSPRVTEAPCPGAPSTASGQTGPQE 540
481 LNKRVQDLSAGGQSLTDSGPERPPEGPAQAPSSPRVTEAPCPGAPSTASGQTGPQE 540

541 PTSARA 546
541 PTSARA 546

RESULT 5
US-10-616-187-44
; Sequence 44, Application US/10616187
; Publication No. US20040013668A1
; GENERAL INFORMATION:
; APPLICANT: Lees, Ann M.
; APPLICANT: Lees, Robert S.
; APPLICANT: Law, Simon W.
; APPLICANT: Arjona, Anibal A.
; TITLE OF INVENTION: NOVEL LOW DENSITY LIPOPROTEIN BINDING
; TITLE OF INVENTION: PROTEINS AND THEIR USE IN DIAGNOSING AND TREATING
; FILE REFERENCE: 10797-004001
; CURRENT APPLICATION NUMBER: US/10/616,187
; PRIOR FILING DATE: 2003-07-09
; PRIOR APPLICATION NUMBER: US/09/616,289
; PRIOR FILING DATE: 2000-07-14
; PRIOR APPLICATION NUMBER: US 09/517,849
; PRIOR FILING DATE: 2000-03-02
; PRIOR APPLICATION NUMBER: US 08/979,608

; ORGANISM: Homo sapiens
US-10-023-529-44

Query Match 100.0%; Score 546; DB 13; Length 546;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 546; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

1 MKNQDKKNGAAGKSNPKSPGQPEAGPEGAQERPSQAAPAVEAGPGSSQAAPKPEGAQA 60
1 MKNQDKKNGAAGKSNPKSPGQPEAGPEGAQERPSQAAPAVEAGPGSSQAAPKPEGAQA 60

61 RTAQSGALRDVSEELSRQLEDILSTYCVDNNGQGGEDGAQGEPAPEPADEAKSRRTYVARN 120
61 RTAQSGALRDVSEELSRQLEDILSTYCVDNNGQGGEDGAQGEPAPEPADEAKSRRTYVARN 120

121 GEPEPTPVNGEKEPSKGDPTNTEIRQSDVEGDRDHRPQEKKAKGLGKEITLLMOTLN 180
121 GEPEPTPVNGEKEPSKGDPTNTEIRQSDVEGDRDHRPQEKKAKGLGKEITLLMOTLN 180

181 TLSTPEKLAALCKKYAELLBEHRNSQKQMLKKQKQSQLVQEKDHLRGEHSKAVLARSK 240
181 TLSTPEKLAALCKKYAELLBEHRNSQKQMLKKQKQSQLVQEKDHLRGEHSKAVLARSK 240

241 LESLCELQHNRSLSKEGVQVAREEKEKKEVTSHFQVTLNDIQLQMEQHNRNSKLQ 300
241 LESLCELQHNRSLSKEGVQVAREEKEKKEVTSHFQVTLNDIQLQMEQHNRNSKLQ 300

301 ENMELAEKLKLIQEVYELREEHIDKVFKHQDLQQLVDAKLQQAQEWLKEAEERHOREKD 360
301 ENMELAEKLKLIQEVYELREEHIDKVFKHQDLQQLVDAKLQQAQEWLKEAEERHOREKD 360

361 FLKKEAVESQRMCELMKQOETHLKOOLALYTEKPEFQNTLSKSSSEVFTTFKQEMEKMTK 420
361 FLKKEAVESQRMCELMKQOETHLKOOLALYTEKPEFQNTLSKSSSEVFTTFKQEMEKMTK 420

421 KIKKLEKTTMYRSRWESSNKALLEMAEKTVDKELQVGIQRLKLCRALQTERND 480
421 KIKKLEKTTMYRSRWESSNKALLEMAEKTVDKELQVGIQRLKLCRALQTERND 480

481 LNKRVQDLSAGGQSLTDSGPERPPEGPAQAPSSPRVTEAPCPGAPSTASGQTGPQE 540
481 LNKRVQDLSAGGQSLTDSGPERPPEGPAQAPSSPRVTEAPCPGAPSTASGQTGPQE 540

541 PTSARA 546
541 PTSARA 546

RESULT 4
US-10-023-523-44
; Sequence 44, Application US/10023523
; Publication No. US20020152485A1
; GENERAL INFORMATION:
; APPLICANT: Lees, Ann M.
; APPLICANT: Law, Simon W.
; APPLICANT: Arjona, Anibal A.
; TITLE OF INVENTION: NOVEL LOW DENSITY LIPOPROTEIN BINDING
; TITLE OF INVENTION: PROTEINS AND THEIR USE IN DIAGNOSING AND TREATING
; FILE REFERENCE: 10797-004001
; CURRENT APPLICATION NUMBER: US/10/023,523
; PRIOR FILING DATE: 2001-12-17
; PRIOR APPLICATION NUMBER: US/09/616,289
; PRIOR FILING DATE: 2000-07-14
; PRIOR APPLICATION NUMBER: US 09/517,849
; PRIOR FILING DATE: 2000-03-02
; PRIOR APPLICATION NUMBER: US 08/979,608
; PRIOR FILING DATE: 1997-11-26
; PRIOR APPLICATION NUMBER: US 60/031,930
; PRIOR FILING DATE: 1996-11-27
; PRIOR APPLICATION NUMBER: US 60/048,547
; PRIOR FILING DATE: 1997-06-03

NUMBER OF SEQ ID NOS: 2700
SOFTWARE: Custom
SEQ ID NO 2134
LENGTH: 510
TYPE: PRT
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: misc feature
LOCATION: (1)-(510)
OTHER INFORMATION: Xaa = any amino acid or nothing
US-10-276-774-2134

Query Match 87.2%; Score 476; DB 12; Length 510;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 476; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MNQDKKNGAAGKQSPKSSPGQPEAGPQCAQERPSQAAPAVEAGPGSSQAPRKPEGQA 60
DB 12 MNQDKKNGAAGKQSPKSSPGQPEAGPQCAQERPSQAAPAVEAGPGSSQAPRKPEGQA 71
QY 61 RTAQSGALRDVSEELSRQLEDILSTYCVNNQGGPGEDGAQGEPAEPDAEKSRITYARN 120
DB 72 RTAQSGALRDVSEELSRQLEDILSTYCVNNQGGPGEDGAQGEPAEPDAEKSRITYARN 131
QY 121 GSEPTPVVNGKEPSKGDPTNTEIRQSDVGDHRRPOBKKAAGLKEITLLMOTLN 180
DB 132 GSEPTPVVNGKEPSKGDPTNTEIRQSDVGDHRRPOBKKAAGLKEITLLMOTLN 191
QY 181 TLSTPEKLAALCKKYAELLSEHRNSQKMLLQKQSQVLQVQKDLHGRHSHKAVLARSK 240
DB 192 TLSTPEKLAALCKKYAELLSEHRNSQKMLLQKQSQVLQVQKDLHGRHSHKAVLARSK 251
QY 241 LESLCRLQHRNRLKKEGVQVQARBEERKEKVTSHFQVTLNDIQLQMEQHRNRSKLQ 300
DB 252 LESLCRLQHRNRLKKEGVQVQARBEERKEKVTSHFQVTLNDIQLQMEQHRNRSKLQ 311
QY 301 ENMELAEKLLIQQYELREEDIDKVFHKLQQLQVDAKLQQAQEMLEAEERHOREKD 360
DB 312 ENMELAEKLLIQQYELREEDIDKVFHKLQQLQVDAKLQQAQEMLEAEERHOREKD 371
QY 361 PLLKEAVESQRMCELMKQOETHLKOALALYTKPEEFQNTLSKSEVFTTFKQEMEKMTK 420
DB 372 PLLKEAVESQRMCELMKQOETHLKOALALYTKPEEFQNTLSKSEVFTTFKQEMEKMTK 431
QY 421 KIKLEKETTYRWRWSSNKALLEMAEKTVDKLEGLQVKTORLEKLCRALQ 476
DB 432 KIKLEKETTYRWRWSSNKALLEMAEKTVDKLEGLQVKTORLEKLCRALQ 487

RESULT 7

US-09-962-055-8

Sequence 8, Application US/09962055

Patent No. US30020052033A1

GENERAL INFORMATION:

APPLICANT: Lees, Ann M. S.

Lees, Robert S.

Law, Simon W.

Arjona, Anibal A.

TITLE OF INVENTION: NOVEL LOW DENSITY LIPOPROTEIN

BINDING PROTEINS AND THEIR USES IN DIAGNOSING AND

TREATING ATHEROSCLEROSIS

NUMBER OF SEQUENCES: 42

CORRESPONDENCE ADDRESS:

ADDRESSEE: Fish & Richardson P.C.

STREET: 225 Franklin Street

CITY: Boston

STATE: MA

COUNTRY: USA

ZIP: 02110-2804

COMPUTER READABLE FORM:

MEDIUM TYPE: Diskette

COMPUTER: IBM Compatible

OPERATING SYSTEM: DOS

PRIOR FILING DATE: 1997-11-26
PRIOR APPLICATION NUMBER: US 60/031,930
PRIOR FILING DATE: 1996-11-27
PRIOR APPLICATION NUMBER: US 60/048,547
PRIOR FILING DATE: 1997-06-03
NUMBER OF SEQ ID NOS: 53
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 44
LENGTH: 546
TYPE: PRT
ORGANISM: Homo sapiens
US-10-616-187-44

Query Match 100.0%; Score 546; DB 15; Length 546;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 546; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MNQDKKNGAAGKQSPKSSPGQPEAGPQCAQERPSQAAPAVEAGPGSSQAPRKPEGQA 60
DB 1 MNQDKKNGAAGKQSPKSSPGQPEAGPQCAQERPSQAAPAVEAGPGSSQAPRKPEGQA 60
QY 61 RTAQSGALRDVSEELSRQLEDILSTYCVNNQGGPGEDGAQGEPAEPDAEKSRITYARN 120
DB 61 RTAQSGALRDVSEELSRQLEDILSTYCVNNQGGPGEDGAQGEPAEPDAEKSRITYARN 120
QY 121 GSEPTPVVNGKEPSKGDPTNTEIRQSDVGDHRRPOBKKAAGLKEITLLMOTLN 180
DB 121 GSEPTPVVNGKEPSKGDPTNTEIRQSDVGDHRRPOBKKAAGLKEITLLMOTLN 180
QY 181 TLSTPEKLAALCKKYAELLSEHRNSQKMLLQKQSQVLQVQKDLHGRHSHKAVLARSK 240
DB 181 TLSTPEKLAALCKKYAELLSEHRNSQKMLLQKQSQVLQVQKDLHGRHSHKAVLARSK 240
QY 241 LESLCRLQHRNRLKKEGVQVQARBEERKEKVTSHFQVTLNDIQLQMEQHRNRSKLQ 300
DB 241 LESLCRLQHRNRLKKEGVQVQARBEERKEKVTSHFQVTLNDIQLQMEQHRNRSKLQ 300
QY 301 ENMELAEKLLIQQYELREEDIDKVFHKLQQLQVDAKLQQAQEMLEAEERHOREKD 360
DB 301 ENMELAEKLLIQQYELREEDIDKVFHKLQQLQVDAKLQQAQEMLEAEERHOREKD 360
QY 361 PLLKEAVESQRMCELMKQOETHLKOALALYTKPEEFQNTLSKSEVFTTFKQEMEKMTK 420
DB 361 PLLKEAVESQRMCELMKQOETHLKOALALYTKPEEFQNTLSKSEVFTTFKQEMEKMTK 420
QY 421 KIKLEKETTYRWRWSSNKALLEMAEKTVDKLEGLQVKTORLEKLCRALQ 480
DB 421 KIKLEKETTYRWRWSSNKALLEMAEKTVDKLEGLQVKTORLEKLCRALQ 480
QY 481 LNKRVQDLSAGQGLSDSGPERRPFGAQAQSSPRVTAPCPYPGAPSTEASQQTGPQE 540
DB 481 LNKRVQDLSAGQGLSDSGPERRPFGAQAQSSPRVTAPCPYPGAPSTEASQQTGPQE 540
QY 541 PTSARA 546
DB 541 PTSARA 546

RESULT 6

US-10-276-774-2134

Sequence 2134, Application US/10276774

Publication No. US20040053245A1

GENERAL INFORMATION:

APPLICANT: Hyseq, Inc.

APPLICANT: Tang, Y. Tom et al

TITLE OF INVENTION: No. US20040053245A1el Nucleic Acids and Polypeptides

FILE REFERENCE: 21272-030

CURRENT APPLICATION NUMBER: US/10/276,774

CURRENT FILING DATE: 2002-11-18

PRIOR APPLICATION NUMBER: 09/560,875

PRIOR FILING DATE: 2000-04-27

PRIOR APPLICATION NUMBER: 09/496,914

PRIOR FILING DATE: 2000-02-03

SOFTWARE: FastSeq for Windows Version 2.0

;;
;; CURRENT APPLICATION DATA:
;; FILING DATE: 24-Sep-2001
;; PRIOR APPLICATION DATA:
;; APPLICATION NUMBER: 08/979,608
;; FILING DATE: 26-Nov-1997
;; APPLICATION NUMBER: US 60/031,930
;; FILING DATE: 27-Nov-1996
;; ATTORNEY/AGENT INFORMATION:
;; NAME: Myers, Louis
;; REGISTRATION NUMBER: 35,965
;; REFERENCE/DOCKET NUMBER: 10797-002001 (formerly 3983/59818)
;; TELECOMMUNICATION INFORMATION:
;; TELEPHONE: 617/542-5070
;; TELEFAX: 617/542-8906
;; INFORMATION FOR SEQ ID NO: 8:
;; SEQUENCE CHARACTERISTICS:
;; LENGTH: 530 amino acids
;; TYPE: amino acid
;; TOPOLOGY: linear
;; MOLECULE TYPE: protein
;; SEQUENCE DESCRIPTION: SEQ ID NO: 8:
US-09-962-055-8

Query Match 78.6%; Score 429; DB 9; Length 530;
Best Local Similarity 99.8%; Pred. No. 0;
Matches 529; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 17 KSPGQPEAGPEGAQRPQAAPAVAEAGPGSSQAPRKEGAQARTAGSALRDVSEELS 76
DB 1 KSPGQPEAGPEGAQRPQAAPAVAEAGPGSSQAPRKEGAQARTAGSALRDVSEELS 60
QY 77 RQLEDILSTYCVNNQGGGEGAGGEPAPEDAEKSRITYVARNGEPEPTPVVNGEKEPS 136
DB 61 RQLEDILSTYCVNNQGGGEGAGGEPAPEDAEKSRITYVARNGEPEPTPVVNGEKEPS 120
QY 137 KGPNTTEIRQSDVGDVDRHRRPQEKKAAGLKGKGLTLLMOTLNTLSTPEEKLAALCKKY 196
DB 121 KGPNTTEIRQSDVGDVDRHRRPQEKKAAGLKGKGLTLLMOTLNTLSTPEEKLAALCKKY 180
QY 197 AELLEHRNSQOMKLLQKQSQOLVQEKDHLRGHSHKAVLARSKLESICREIQRNRSK 256
DB 181 AELLEHRNSQOMKLLQKQSQOLVQEKDHLRGHSHKAVLARSKLESICREIQRNRSK 240
QY 257 EGVQVAREEERKEKVTSHFQVTLNDIQLQMEQHNRNSKLRQENMELAEKLLIQY 316
DB 241 EGVQVAREEERKEKVTSHFQVTLNDIQLQMEQHNRNSKLRQENMELAEKLLIQY 300
QY 317 ELREEHIDKVPFKHDLQQLVDKQLQQAQEMLKEAERHOREKDFLLKEAVESQRMCELM 376
DB 301 ELREEHIDKVPFKHDLQQLVDKQLQQAQEMLKEAERHOREKDFLLKEAVESQRMCELM 360
QY 377 KQOETHLQKQALALYTEKPEFQNTLSKSEVFTTFKQEMKMTKKIKKLEKETTMYRSRW 436
DB 361 KQOETHLQKQALALYTEKPEFQNTLSKSEVFTTFKQEMKMTKKIKKLEKETTMYRSRW 420
QY 437 ESSNKALLEMAEKTVDKLEGLQVKIQRLKLCRALQTERNDLNKRVQDLSAGQGSGL 496
DB 421 ESSNKALLEMAEKTVDKLEGLQVKIQRLKLCRALQTERNDLNKRVQDLSAGQGSGL 480
QY 497 TDSGPERPEGGAQAPSSPRVTEAPCPGAPSTASGQTGPQETSARA 546
DB 481 TDSGPERPEGGAQAPSSPRVTEAPCPGAPSTASGQTGPQETSARA 530

RESULT 8

US-09-976-740-8
; Sequence 8, Application US/09976740
; Publication No. US20020194633A1
; GENERAL INFORMATION:
; APPLICANT: Lees, Ann M.
; APPLICANT: Lees, Robert S.
; APPLICANT: Law, Simon W.

;;
;; APPLICANT: Law, Simon W.
;; APPLICANT: Arjona, Anibal A.
;; TITLE OF INVENTION: NOVEL LOW DENSITY LIPOPROTEIN BINDING
;; TITLE OF INVENTION: PROTEINS AND THEIR USE IN DIAGNOSING AND TREATING
;; TITLE OF INVENTION: ATHEROSCLEROSIS
;; FILE REFERENCE: 10797-004001
;; CURRENT APPLICATION NUMBER: US/09/976,740
;; CURRENT FILING DATE: 2001-10-12
;; PRIOR APPLICATION NUMBER: 09/616,289
;; PRIOR FILING DATE: 2000-07-14
;; PRIOR APPLICATION NUMBER: US 08/979,608
;; PRIOR FILING DATE: 1997-11-26
;; PRIOR APPLICATION NUMBER: US 60/031,930
;; PRIOR FILING DATE: 1996-11-27
;; PRIOR APPLICATION NUMBER: US 60/048,547
;; PRIOR FILING DATE: 1997-06-03
;; NUMBER OF SEQ ID NOS: 53
;; SOFTWARE: FastSeq for Windows Version 4.0
;; SEQ ID NO 8
;; LENGTH: 530
;; TYPE: PRT
;; ORGANISM: Homo sapiens
US-09-976-740-8

Query Match 78.6%; Score 429; DB 9; Length 530;

Best Local Similarity 99.8%; Pred. No. 0;
Matches 529; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 17 KSPGQPEAGPEGAQRPQAAPAVAEAGPGSSQAPRKEGAQARTAGSALRDVSEELS 76
DB 1 KSPGQPEAGPEGAQRPQAAPAVAEAGPGSSQAPRKEGAQARTAGSALRDVSEELS 60
QY 77 RQLEDILSTYCVNNQGGGEGAGGEPAPEDAEKSRITYVARNGEPEPTPVVNGEKEPS 136
DB 61 RQLEDILSTYCVNNQGGGEGAGGEPAPEDAEKSRITYVARNGEPEPTPVVNGEKEPS 120
QY 137 KGPNTTEIRQSDVGDVDRHRRPQEKKAAGLKGKGLTLLMOTLNTLSTPEEKLAALCKKY 196
DB 121 KGPNTTEIRQSDVGDVDRHRRPQEKKAAGLKGKGLTLLMOTLNTLSTPEEKLAALCKKY 180
QY 197 AELLEHRNSQOMKLLQKQSQOLVQEKDHLRGHSHKAVLARSKLESICREIQRNRSK 256
DB 181 AELLEHRNSQOMKLLQKQSQOLVQEKDHLRGHSHKAVLARSKLESICREIQRNRSK 240
QY 257 EGVQVAREEERKEKVTSHFQVTLNDIQLQMEQHNRNSKLRQENMELAEKLLIQY 316
DB 241 EGVQVAREEERKEKVTSHFQVTLNDIQLQMEQHNRNSKLRQENMELAEKLLIQY 300
QY 317 ELREEHIDKVPFKHDLQQLVDKQLQQAQEMLKEAERHOREKDFLLKEAVESQRMCELM 376
DB 301 ELREEHIDKVPFKHDLQQLVDKQLQQAQEMLKEAERHOREKDFLLKEAVESQRMCELM 360
QY 377 KQOETHLQKQALALYTEKPEFQNTLSKSEVFTTFKQEMKMTKKIKKLEKETTMYRSRW 436
DB 361 KQOETHLQKQALALYTEKPEFQNTLSKSEVFTTFKQEMKMTKKIKKLEKETTMYRSRW 420
QY 437 ESSNKALLEMAEKTVDKLEGLQVKIQRLKLCRALQTERNDLNKRVQDLSAGQGSGL 496
DB 421 ESSNKALLEMAEKTVDKLEGLQVKIQRLKLCRALQTERNDLNKRVQDLSAGQGSGL 480
QY 497 TDSGPERPEGGAQAPSSPRVTEAPCPGAPSTASGQTGPQETSARA 546
DB 481 TDSGPERPEGGAQAPSSPRVTEAPCPGAPSTASGQTGPQETSARA 530

RESULT 9

US-10-671-242-8
; Sequence 8, Application US/10671242
; Publication No. US20040040049A1
; GENERAL INFORMATION:
; APPLICANT: Lees, Ann M.
; APPLICANT: Lees, Robert S.
; APPLICANT: Law, Simon W.

```

1  APPLICANT: Law, Simon W.
2  APPLICANT: Arjona, Anibal A.
3  TITLE OF INVENTION: NOVEL LOW DENSITY LIPOPROTEIN BINDING
4  TITLE OF INVENTION: PROTEINS AND THEIR USE IN DIAGNOSING AND TREATING
5  TITLE OF INVENTION: ATHEROSCLEROSIS
6  FILS REFERENCE: 10797-004001
7  CURRENT APPLICATION NUMBER: US/10/023,529
8  CURRENT FILING DATE: 2001-12-17
9  PRIOR APPLICATION NUMBER: 09/616,289
10 PRIOR FILING DATE: 2000-07-14
11 PRIOR APPLICATION NUMBER: US 09/517,849
12 PRIOR FILING DATE: 2000-03-02
13 PRIOR APPLICATION NUMBER: US 08/979,608
14 PRIOR FILING DATE: 1997-11-26
15 PRIOR APPLICATION NUMBER: US 60/031,930
16 PRIOR FILING DATE: 1996-11-27
17 PRIOR APPLICATION NUMBER: US 60/048,547
18 PRIOR FILING DATE: 1997-06-03
19 NUMBER OF SEQ ID NOS: 53
20 SOFTWARE: fastseq for Windows Version 4.0
21 SEQ ID NO 8
22 LENGTH: 530
23 TYPE: PRT
24 ORGANISM: Homo sapiens
25 US-10-023-529-8

```

	Query Match	78.6%;	Score 429;	DB 13;	Length 530;	
	Best Local Similarity	99.8%;	Pred. No. 0;			
	Matches 529;	Conservative	0;	Mismatches	1;	Indels
						0;
Qy	17	KSPGQPEAGPEGAQERPSQAAPAVAEAGPGSSQAAPRKPEGGAQARTAQSGALRDVSEELS	76			
Db	1	KSPGQPEAGPEGAQERPSQAAPAVAEAGPGSSQAAPRKPEGGAQARTAQSGALRDVSEELS	60			
Qy	77	RLQEDILSTCYDNNGGGEGDGAQGEPAPEDAEKSRTYVAANGPEPTPPVVGKEPS	136			
Db	61	RLQEDILSTCYDNNGGGEGDGAQGEPAPEDAEKSRTYVAANGPEPTPPVVGKEPS	120			
Qy	137	KGDPNTEETIQSDVGDGRHRRPOEKKKAKGLGKEITLLMQTLNTLSTPEKLAALCKKY	196			
Db	121	KGDPNTEETIQSDVGDGRHRRPOEKKKAKGLGKEITLLMQTLNTLSTPEKLAALCKKY	180			
Qy	197	AELLEEHRNSQKQMLLQKKQSQLVQEKDHLRGEHSKAVLARSKLSLCELQRHNRSLX	256			
Db	181	AELLEEHRNSQKQMLLQKKQSQLVQEKDHLRGEHSKAVLARSKLSLCELQRHNRSLX	240			
Qy	257	EGVQVAREEEERKEVTSHPQVTLNDIQLQMECHNRNSKLQENNELAERLKLLEQY	316			
Db	241	EGVQVAREEEERKEVTSHPQVTLNDIQLQMECHNRNSKLQENNELAERLKLLEQY	300			
Qy	317	ELREEHIDKVPFKHKLQQLQVDAKLAQAQEWLKEABERHOREDFLKEAVESQRMCELM	375			
Db	301	ELREEHIDKVPFKHKLQQLQVDAKLAQAQEWLKEABERHOREDFLKEAVESQRMCELM	360			
Qy	377	KQETHLKKQALALYTKPEPFQNTLSKSSVPTTFQEMEKMTYKIKKLKEITMYESRW	436			
Db	361	KQETHLKKQALALYTKPEPFQNTLSKSSVPTTFQEMEKMTYKIKKLKEITMYESRW	420			
Qy	437	ESNNKALLEWABEKTVRDKELEGLOVKIQRLEKLCRALQTERNDLNKRVODLSAGGQSL	496			
Db	421	ESNNKALLEWABEKTVRDKELEGLOVKIQRLEKLCRALQTERNDLNKRVODLSAGGQSL	480			
Qy	497	TDGSPERRPBGGAQAPSSPRVTAPCYCAPSTASGGTGPOEPTSARA	546			
Db	481	TDGSPERRPBGGAQAPSSPRVTAPCYCAPSTASGGTGPOEPTSARA	530			

RESULT 11
US-10-023-523-8
; Sequence 8, Application US/10023523
; Publication No. US20020152485A1
; GENERAL INFORMATION:
; APPLICANT: Lees, Ann M.

; APPLICANT: Lees, Robert S.
; APPLICANT: Law, Simon W.
; APPLICANT: Arjona, Anibal A.
; TITLE OF INVENTION: NOVEL LOW DENSITY LIPOPROTEIN BINDING
; TITLE OF INVENTION: PROTEINS AND THEIR USE IN DIAGNOSING AND TREATING
; TITLE OF INVENTION: ATHEROSCLEROSIS
; FILE REFERENCE: 10797-004001
; CURRENT APPLICATION NUMBER: US/10/023,523
; CURRENT FILING DATE: 2001-12-17
; PRIOR APPLICATION NUMBER: US/09/616,289
; PRIOR FILING DATE: 2000-07-14
; PRIOR APPLICATION NUMBER: US 09/517,849
; PRIOR FILING DATE: 2000-03-02
; PRIOR APPLICATION NUMBER: US 08/979,608
; PRIOR FILING DATE: 1997-11-26
; PRIOR APPLICATION NUMBER: US 60/031,930
; PRIOR FILING DATE: 1996-11-27
; PRIOR APPLICATION NUMBER: US 60/048,547
; PRIOR FILING DATE: 1997-06-03
; NUMBER OF SEQ ID NOS: 53
; SOFTWARE: PastSeq for Windows Version 4.0
; SEQ ID NO 8
; LENGTH: 530
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-023-523-8

Query Match 78.6%; Score 429; DB 13; Length 530;
Best Local Similarity 99.8%; Pred. No. 0;
Matches 529; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 17 KSSPGQPEAGPEGAQRRPSQAAPVAEAGPGSSQAAPRKEGAQARTAQSGALRDVSEELS 76
Db 1 KSSPGQPEAGPEGAQRRPSQAAPVAEAGPGSSQAAPRKEGAQARTAQSGALRDVSEELS 60
QY 77 RQLEDILSTYCVDNNGGPGEDGAGCEPAEPAEDAEKSTTYVARGNPEPTPVVNGEKEPS 136
Db 61 RQLEDILSTYCVDNNGGPGEDGAGCEPAEPAEDAEKSTTYVARGNPEPTPVVNGEKEPS 120
QY 137 KGPNTTEIRQSDVGDHRRPOEKKKAGLGKEITLLMQTLNTLSTPEEKLAALCKKY 196
Db 121 KGPNTTEIRQSDVGDHRRPOEKKKAGLGKEITLLMQTLNTLSTPEEKLAALCKKY 180
QY 197 AELLLEHNSQOMKLLQKQSQLVQEKDHLRGHNSKAVLARSKLSLCSRLQHRNRSJK 256
Db 181 AELLLEHNSQOMKLLQKQSQLVQEKDHLRGHNSKAVLARSKLSLCSRLQHRNRSJK 240
QY 257 EGVQARBEERKEKVTSHFQVTLNDIQLQMEQHNRNSKLRQENMELAEKLLIEQY 316
Db 241 EGVQARBEERKEKVTSHFQVTLNDIQLQMEQHNRNSKLRQENMELAEKLLIEQY 300
QY 317 ELREEHIDKVFHKDLQOOLVDAKLQQAQEMLKEAEERHOREKDPLLKEAVESQRMCELM 376
Db 301 ELREEHIDKVFHKDLQOOLVDAKLQQAQEMLKEAEERHOREKDPLLKEAVESQRMCELM 360
QY 377 KQETHLKOQALALYTEKEEPONTLSKSSSEVFTTTPKQEMKMTKKIKKLEKTTMYRSW 436
Db 361 KQETHLKOQALALYTEKEEPONTLSKSSSEVFTTTPKQEMKMTKKIKKLEKTTMYRSW 420
QY 437 ESSNKALLEMAEEKTVRDKELEGQVKIQRLEKLCRALQTERNDLNKRVDLSAGQGSGL 496
Db 421 ESSNKALLEMAEEKTVRDKELEGQVKIQRLEKLCRALQTERNDLNKRVDLSAGQGSGL 480
QY 497 TDGSPRRPEPGAQAPSSPRVTEAPCYPGAPSTEASGOTGPOEPTSARA 546
Db 481 TDGSPRRPEPGAQAPSSPRVTEAPCYPGAPSTEASGOTGPOEPTSARA 530

RESULT 12
US-10-616-187-8
; Sequence 8, Application US/10616187
; Publication No. US20040013668A1
; GENERAL INFORMATION:

; APPLICANT: Lees, Ann M.
; APPLICANT: Lees, Robert S.
; APPLICANT: Law, Simon W.
; APPLICANT: Arjona, Anibal A.
; TITLE OF INVENTION: NOVEL LOW DENSITY LIPOPROTEIN BINDING
; TITLE OF INVENTION: PROTEINS AND THEIR USE IN DIAGNOSING AND TREATING
; TITLE OF INVENTION: ATHEROSCLEROSIS
; FILE REFERENCE: 10797-004001
; CURRENT APPLICATION NUMBER: US/10/616,187
; CURRENT FILING DATE: 2003-07-09
; PRIOR APPLICATION NUMBER: US/09/616,289
; PRIOR FILING DATE: 2000-07-14
; PRIOR APPLICATION NUMBER: US 09/517,849
; PRIOR FILING DATE: 2000-03-02
; PRIOR APPLICATION NUMBER: US 08/979,608
; PRIOR FILING DATE: 1997-11-26
; PRIOR APPLICATION NUMBER: US 60/031,930
; PRIOR FILING DATE: 1996-11-27
; PRIOR APPLICATION NUMBER: US 60/048,547
; PRIOR FILING DATE: 1997-06-03
; NUMBER OF SEQ ID NOS: 53
; SOFTWARE: PastSeq for Windows Version 4.0
; SEQ ID NO 8
; LENGTH: 530
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-616-187-8

Query Match 78.6%; Score 429; DB 15; Length 530;
Best Local Similarity 99.8%; Pred. No. 0;
Matches 529; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 17 KSSPGQPEAGPEGAQRRPSQAAPVAEAGPGSSQAAPRKEGAQARTAQSGALRDVSEELS 76
Db 1 KSSPGQPEAGPEGAQRRPSQAAPVAEAGPGSSQAAPRKEGAQARTAQSGALRDVSEELS 60
QY 77 RQLEDILSTYCVDNNGGPGEDGAGCEPAEPAEDAEKSTTYVARGNPEPTPVVNGEKEPS 136
Db 61 RQLEDILSTYCVDNNGGPGEDGAGCEPAEPAEDAEKSTTYVARGNPEPTPVVNGEKEPS 120
QY 137 KGPNTTEIRQSDVGDHRRPOEKKKAGLGKEITLLMQTLNTLSTPEEKLAALCKKY 196
Db 121 KGPNTTEIRQSDVGDHRRPOEKKKAGLGKEITLLMQTLNTLSTPEEKLAALCKKY 180
QY 197 AELLLEHNSQOMKLLQKQSQLVQEKDHLRGHNSKAVLARSKLSLCSRLQHRNRSJK 256
Db 181 AELLLEHNSQOMKLLQKQSQLVQEKDHLRGHNSKAVLARSKLSLCSRLQHRNRSJK 240
QY 257 EGVQARBEERKEKVTSHFQVTLNDIQLQMEQHNRNSKLRQENMELAEKLLIEQY 316
Db 241 EGVQARBEERKEKVTSHFQVTLNDIQLQMEQHNRNSKLRQENMELAEKLLIEQY 300
QY 317 ELREEHIDKVFHKDLQOOLVDAKLQQAQEMLKEAEERHOREKDPLLKEAVESQRMCELM 376
Db 301 ELREEHIDKVFHKDLQOOLVDAKLQQAQEMLKEAEERHOREKDPLLKEAVESQRMCELM 360
QY 377 KQETHLKOQALALYTEKEEPONTLSKSSSEVFTTTPKQEMKMTKKIKKLEKTTMYRSW 436
Db 361 KQETHLKOQALALYTEKEEPONTLSKSSSEVFTTTPKQEMKMTKKIKKLEKTTMYRSW 420
QY 437 ESSNKALLEMAEEKTVRDKELEGQVKIQRLEKLCRALQTERNDLNKRVDLSAGQGSGL 496
Db 421 ESSNKALLEMAEEKTVRDKELEGQVKIQRLEKLCRALQTERNDLNKRVDLSAGQGSGL 480
QY 497 TDGSPRRPEPGAQAPSSPRVTEAPCYPGAPSTEASGOTGPOEPTSARA 546
Db 481 TDGSPRRPEPGAQAPSSPRVTEAPCYPGAPSTEASGOTGPOEPTSARA 530

RESULT 13
US-09-962-055-5
; Sequence 5, Application US/09962055
; Patent No. US2002005203A1

	Matches	172;	Conservative	0;	Mismatches	0;	Indels	0;	Gaps	0;
Qy	280	TLNDIQLQWQCHNERN	SKLRQENMELAE	RLKKLIBQYELREEH	IDKVPFKHDLQOOL	YDA	339			
Db	280	TLNDIQLQWQCHNERN	SKLRQENMELAE	RLKKLIBQYELREEH	IDKVPFKHDLQOOL	YDA	339			
Qy	340	KLQQAQEMLKAEERH	QREKDFLLKEAVES	QRMCELMKQOETHL	KQOALALYTEKFE	EQN	399			
Db	340	KLQQAQEMLKAEERH	QREKDFLLKEAVES	QRMCELMKQOETHL	KQOALALYTEKFE	EQN	399			
Qy	400	TLSSKSEVFTTQEMEK	MTKKIKLEKETTT	MYRSRWSSNKALLE	MEBKT	451				
Db	400	TLSSKSEVFTTQEMEK	MTKKIKLEKETTT	MYRSRWSSNKALLE	MEBKT	451				

Search completed: June 8, 2004, 16:41:32
Job time : 52 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: June 7, 2004, 14:41:31 ; Search time 40 Seconds
(without alignments)
4180.616 Million cell updates/sec

Title: US-10-023-529-8
Perfect score: 530
Sequence: 1 KSSPQPRAGPEGAQERPSQ.....APSTEASQTGPQRPSTARA 530

Scoring table: OLIGO
Gapop 60.0 , Gapext 60.0

Searched: 1017041 seqs, 315518202 residues

Word size : 0
Total number of hits satisfying chosen parameters: 1017041

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Listing first 300 summaries

Database : SPTREML25.*

- 1: sp_archaea.*
- 2: sp_bacteria.*
- 3: sp_fungi.*
- 4: sp_human.*
- 5: sp_invertebrate.*
- 6: sp_mammal.*
- 7: sp_mhc.*
- 8: sp_organelle.*
- 9: sp_phage.*
- 10: sp_plant.*
- 11: sp_rodent.*
- 12: sp_virus.*
- 13: sp_vertibrate.*
- 14: sp_unclassified.*
- 15: sp_rvirus.*
- 16: sp_bacteriap.*
- 17: sp_archaeap.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	429	80.9	546	4 Q86VW3	Q86VW3 homo sapien
2	412	77.7	546	4 Q86T54	Q86T54 homo sapien
3	328	61.9	546	4 Q86T86	Q86T86 homo sapien
4	178	33.6	546	4 Q86T85	Q86T85 homo sapien
5	116	21.9	116	4 Q86Y86	Q86Y86 homo sapien
6	21	4.0	150	11 Q86WJ7	Q86WJ7 mus musculus
7	19	3.6	241	11 Q86P11	Q86P11 mus musculus
8	19	3.6	241	11 Q86P11	Q86P11 mus musculus
9	19	3.6	463	11 Q86UK2	Q86UK2 mus musculus
10	19	3.6	524	11 Q86HNT	Q86HNT mus musculus
11	19	3.6	528	4 Q9NUQ3	Q9NUQ3 homo sapien
12	19	3.6	684	11 Q8VBT1	Q8VBT1 mus musculus
13	17	3.2	186	4 Q9POX1	Q9POX1 homo sapien
14	17	3.2	505	4 Q8X3S2	Q8X3S2 homo sapien
15	17	3.2	676	13 Q81969	Q81969 gallus gall
16	17	3.2	715	4 Q8N3L3	Q8N3L3 homo sapien

17	3.2	718	4	Q86T52	Q86T52 homo sapien
18	2.1	404	10	Q8RXD7	Q8RXD7 arabidopsis
19	1.9	181	4	Q8BZA4	Q8BZA4 homo sapien
20	1.9	190	11	Q8BUX8	Q8BUX8 mus musculus
21	1.7	637	16	Q8DTC1	Q8DTC1 streptococc
22	9	826	5	Q8VPS3	Q8VPS3 streptococc
23	1.7	842	3	Q86WP6	Q86WP6 lentinula e
24	1.5	113	5	Q86I22	Q86I22 dictyosteli
25	8	135	16	Q8PL61	Q8PL61 xanthomonas
26	8	135	16	Q8P9E8	Q8P9E8 xanthomonas
27	8	143	2	P70717	P70717 actinobacil
28	8	195	5	O17194	O17194 caenorhabdi
29	8	206	16	Q83C21	Q83C21 coxiella bu
30	8	224	10	Q8SCX3	Q8SCX3 arabidopsis
31	8	229	16	Q86077	Q86077 mycobacteri
32	8	255	16	Q8R9Z0	Q8R9Z0 thermomanaer
33	8	255	16	Q8R9Z9	Q8R9Z9 pseudomonas
34	8	264	5	Q86LQ1	Q86LQ1 branchiosto
35	8	281	5	O44174	O44174 caenorhabdi
36	8	312	10	Q8SZW2	Q8SZW2 arabidopsis
37	8	324	10	Q8SA04	Q8SA04 oryza sativ
38	8	346	16	Q8BLJ3	Q8BLJ3 oceanobacil
39	8	348	12	Q9QIU7	Q9QIU7 saimirine
40	8	348	12	Q80SG5	Q80SG5 saimirine
41	8	395	2	Q9AES1	Q9AES1 bacteroides
42	8	405	12	Q80BK0	Q80BK0 saimirine
43	8	405	12	Q80SM7	Q80SM7 saimirine
44	8	413	16	Q82DR9	Q82DR9 streptomyce
45	8	434	16	Q9LQW6	Q9LQW6 streptomyce
46	8	436	10	Q84VB3	Q84VB3 oryza sativ
47	8	439	12	Q80BH6	Q80BH6 saimirine
48	8	450	16	Q9ZLG8	Q9ZLG8 helicobacte
49	8	451	3	Q8J2S6	Q8J2S6 gibberella
50	8	484	16	Q8NMB0	Q8NMB0 corynebacte
51	8	501	16	Q8PYI9	Q8PYI9 xanthomonas
52	8	520	12	Q9QIU8	Q9QIU8 saimirine
53	8	530	11	Q8VCQ8	Q8VCQ8 mus musculu
54	8	539	16	Q8P3J1	Q8P3J1 xanthomonas
55	8	558	10	Q93ZM4	Q93ZM4 arabidopsis
56	8	568	17	Q97WX9	Q97WX9 sulfolobus
57	8	609	16	Q8DJC0	Q8DJC0 synecococc
58	8	649	16	Q98P27	Q98P27 rhizobium l
59	8	649	16	Q98P27	Q98P27 rhizobium m
60	8	928	10	O48931	O48931 arabidopsis
61	8	929	10	Q8P156	Q8P156 arabidopsis
62	8	944	4	Q9UP82	Q9UP82 homo sapien
63	8	952	10	Q9M2Z6	Q9M2Z6 arabidopsis
64	8	952	10	Q9SXJ7	Q9SXJ7 arabidopsis
65	8	1021	16	Q7V110	Q7V110 helicobacte
66	8	1045	4	Q76037	Q76037 homo sapien
67	8	1101	16	Q7UYN4	Q7UYN4 rhodospirill
68	8	1107	4	Q9H4D6	Q9H4D6 homo sapien
69	8	1157	4	Q9GZM2	Q9GZM2 homo sapien
70	8	1187	4	Q9GZV6	Q9GZV6 homo sapien
71	8	1199	4	Q9H487	Q9H487 homo sapien
72	8	1201	4	Q9H486	Q9H486 homo sapien
73	8	1214	4	Q9H485	Q9H485 homo sapien
74	8	1215	4	Q9H484	Q9H484 homo sapien
75	8	1256	4	Q9H483	Q9H483 homo sapien
76	8	1290	16	Q8PKZ7	Q8PKZ7 xanthomonas
77	8	1827	4	Q9H482	Q9H482 homo sapien
78	8	1902	16	Q92JF8	Q92JF8 rickettsia
79	8	2117	4	Q9H4D8	Q9H4D8 homo sapien
80	8	2167	4	Q9H481	Q9H481 homo sapien
81	8	2169	4	Q9NY09	Q9NY09 homo sapien
82	7	23	5	Q26676	Q26676 tethya aura
83	7	23	12	Q91K25	Q91K25 heparitis c
84	7	42	16	Q8FRQ0	Q8FRQ0 escherichia
85	7	42	12	Q72301	Q72301 heparitis c
86	7	60	16	Q8X2X0	Q8X2X0 escherichia
87	7	62	16	Q8YQ30	Q8YQ30 anabaena sp
88	7	69	12	Q72305	Q72305 heparitis c
89	7	70	11	Q8VCP1	Q8VCP1 mus musculu

90	7	1.3	74	12	Q9W1K7	Q9w1k7 hepatitis c	163	7	1.3	172	17	O28639	O28639 archaeoglob
91	7	1.3	75	12	Q9W9I6	Q9w9i6 hepatitis c	164	7	1.3	173	4	Q9Y5A0	Q9y5a0 homo sapien
92	7	1.3	76	16	Q9Z8F1	Q9z8f1 listeria in	165	7	1.3	176	5	Q9Y5A0	Q9y5a0 schistosoma
93	7	1.3	77	16	Q9Z8F1	Q9z8f1 listeria in	166	7	1.3	177	5	O76627	O76627 caenorhabdi
94	7	1.3	78	16	Q9RSW4	Q9rsW4 deinococcus	167	7	1.3	177	16	O8G322	O8g322 bifidobacte
95	7	1.3	79	16	Q9RSW4	Q9rsW4 deinococcus	168	7	1.3	179	10	O23356	O23356 arabidopsis
96	7	1.3	80	16	Q9SL13	Q9sl13 shigella fl	169	7	1.3	180	16	O8YF7H	O8yF7H brucella me
97	7	1.3	81	16	Q9W811	Q9w811 pseudomonas	170	7	1.3	181	16	O50914	O50914 borrelia bu
98	7	1.3	82	2	O66133	O66133 acidiphiliu	171	7	1.3	181	16	O50914	O50914 borrelia bu
99	7	1.3	83	16	O92J77	O92j77 rhizobium m	172	7	1.3	186	2	Q9AEN4	Q9aen4 burkholderi
100	7	1.3	84	16	O92J77	O92j77 rhizobium m	173	7	1.3	187	16	Q92XC5	Q92xc5 rhizobium m
101	7	1.3	85	2	O66135	O66135 acidiphiliu	174	7	1.3	190	1	P94913	P94913 methanosarc
102	7	1.3	86	10	Q8L5Q4	Q8l5q4 cicier ariet	175	7	1.3	190	1	P94913	P94913 methanosarc
103	7	1.3	87	10	Q8WR28	Q8wr28 anophelies g	176	7	1.3	193	17	Q96XQ6	Q96xq6 sulfolobus
104	7	1.3	88	16	Q83KC6	Q83kc6 shigella fl	177	7	1.3	193	17	Q96XQ6	Q96xq6 sulfolobus
105	7	1.3	89	3	P79038	P79038 emericeila	178	7	1.3	194	12	O81843	O81843 hepatitis d
106	7	1.3	90	15	Q9Q4Y7	Q9q4y7 human immun	179	7	1.3	194	12	O91DH7	O91dh7 hepatitis d
107	7	1.3	91	10	Q8H5T1	Q8h5t1 oryza sativ	180	7	1.3	194	12	O91DH7	O91dh7 hepatitis d
108	7	1.3	92	10	Q8H5T1	Q8h5t1 oryza sativ	181	7	1.3	194	12	O91DH7	O91dh7 hepatitis d
109	7	1.3	93	12	O72304	O72304 hepatitis c	182	7	1.3	195	12	O92E26	O92e26 hepatitis d
110	7	1.3	94	12	O8C992	O8c992 mmu muscucu	183	7	1.3	195	12	O92E26	O92e26 hepatitis d
111	7	1.3	95	12	O72300	O72300 hepatitis c	184	7	1.3	195	12	O88972	O88972 hepatitis d
112	7	1.3	96	12	O72302	O72302 hepatitis c	185	7	1.3	195	12	O88972	O88972 hepatitis d
113	7	1.3	97	12	O72299	O72299 hepatitis c	186	7	1.3	195	12	O88972	O88972 hepatitis d
114	7	1.3	98	16	O7TV63	O7tv63 prochloroco	187	7	1.3	200	10	O7X7T1	O7x7t1 oryza sativ
115	7	1.3	99	12	O72297	O72297 hepatitis c	188	7	1.3	201	16	O92X75	O92x75 rhizobium m
116	7	1.3	100	12	Q9W800	Q9w800 hepatitis c	189	7	1.3	203	10	O7XK48	O7xk48 oryza sativ
117	7	1.3	101	12	Q9W879	Q9w879 hepatitis c	190	7	1.3	205	10	O94A87	O94a87 arabidopsis
118	7	1.3	102	12	Q9W878	Q9w878 hepatitis c	191	7	1.3	208	2	O68819	O68819 brucella me
119	7	1.3	103	10	Q949K5	Q949k5 lycopersico	192	7	1.3	208	16	O8G2D9	O8g2d9 brucella su
120	7	1.3	104	10	Q9M4H2	Q9m4h2 vitis vinif	193	7	1.3	208	16	O8G2D9	O8g2d9 brucella su
121	7	1.3	105	16	O81F87	O81f87 bacillus ce	194	7	1.3	211	2	Q8VUW2	Q8vuw2 staphylococ
122	7	1.3	106	2	Q9XCD3	Q9xcd3 thermoanos	195	7	1.3	212	16	Q93CM8	Q93cm8 salmonella
123	7	1.3	107	4	Q8N8Z7	Q8n8z7 homo sapien	196	7	1.3	212	16	Q93CM8	Q93cm8 salmonella
124	7	1.3	108	5	Q81TB3	Q81tb3 aequipecten	197	7	1.3	214	12	O9PX77	O9px77 hepatitis d
125	7	1.3	109	2	Q8V8I9	Q8v8i9 shigella fl	198	7	1.3	214	12	O9PX77	O9px77 hepatitis d
126	7	1.3	110	16	Q89A73	Q89a73 buchnera ap	199	7	1.3	214	12	O9PZU5	O9pzU5 hepatitis d
127	7	1.3	111	16	Q81ZM3	Q81zm3 shigella fl	200	7	1.3	214	12	O81842	O81842 hepatitis d
128	7	1.3	112	16	Q81ZM2	Q81zm2 shigella fl	201	7	1.3	214	12	O81842	O81842 hepatitis d
129	7	1.3	113	16	Q81ZM2	Q81zm2 shigella fl	202	7	1.3	214	12	O9PZV4	O9pzv4 hepatitis d
130	7	1.3	114	16	Q81ZM2	Q81zm2 shigella fl	203	7	1.3	214	12	O9PZV4	O9pzv4 hepatitis d
131	7	1.3	115	16	Q7UBX8	Q7ubx8 shigella fl	204	7	1.3	214	12	O81835	O81835 hepatitis d
132	7	1.3	116	16	O7TTG1	O7ttg1 shigella fl	205	7	1.3	214	12	O9PZV1	O9pzv1 hepatitis d
133	7	1.3	117	5	Q95T28	Q95t28 drosophila	206	7	1.3	214	12	O68969	O68969 hepatitis d
134	7	1.3	118	16	Q97Z73	Q97z73 sulfolobus	207	7	1.3	214	12	O9PZV0	O9pzv0 hepatitis d
135	7	1.3	119	16	Q8F9K0	Q8f9k0 leptospira	208	7	1.3	214	12	O9PZV2	O9pzv2 hepatitis d
136	7	1.3	120	16	Q9LMV6	Q9lmv6 arabidopsis	209	7	1.3	214	12	O9PZV2	O9pzv2 hepatitis d
137	7	1.3	121	6	Q9MTY7	Q9myy7 oryctolagus	210	7	1.3	214	12	O9PZV6	O9pzv6 hepatitis d
138	7	1.3	122	6	Q9MTY6	Q9myy6 sus scrofa	211	7	1.3	214	12	O9PZV6	O9pzv6 hepatitis d
139	7	1.3	123	13	Q9F8H7	Q9f8h7 gallus gall	212	7	1.3	214	12	O9PZU3	O9pzU3 hepatitis d
140	7	1.3	124	2	O52591	O52591 mycobacteri	213	7	1.3	214	12	O9PZU3	O9pzU3 hepatitis d
141	7	1.3	125	10	Q7XML1	Q7xml1 oryza sativ	214	7	1.3	214	12	O8BC70	O8bc70 hepatitis d
142	7	1.3	126	2	Q9AFN4	Q9afn4 shigella fl	215	7	1.3	214	12	O8BC70	O8bc70 hepatitis d
143	7	1.3	127	2	Q7X4U9	Q7x4u9 streptococc	216	7	1.3	214	12	O80JX1	O80jx1 hepatitis d
144	7	1.3	128	2	Q7X4U8	Q7x4u8 streptococc	217	7	1.3	215	16	O88V70	O88v70 lactobacill
145	7	1.3	129	2	Q7X4U5	Q7x4u5 streptococc	218	7	1.3	220	12	O9PZC4	O9pzc4 st. louis e
146	7	1.3	130	2	Q7X4U4	Q7x4u4 streptococc	219	7	1.3	220	12	O9PZC4	O9pzc4 st. louis e
147	7	1.3	131	2	Q7X4U2	Q7x4u2 streptococc	220	7	1.3	221	2	Q847E0	Q847e0 pseudomonas
148	7	1.3	132	2	Q7X4U0	Q7x4u0 streptococc	221	7	1.3	221	14	P97973	P97973 unidentified
149	7	1.3	133	2	Q7X4T8	Q7x4t8 streptococc	222	7	1.3	226	2	Q8WPF3	Q8wvf3 pseudomonas
150	7	1.3	134	2	Q7X4T6	Q7x4t6 streptococc	223	7	1.3	227	16	Q7UCD8	Q7ucd8 shigella fl
151	7	1.3	135	2	Q7X4T4	Q7x4t4 streptococc	224	7	1.3	227	16	Q8U3V9	Q8u3v9 pyrococcus
152	7	1.3	136	2	Q7WRH1	Q7wrh1 streptococc	225	7	1.3	229	16	O8X4Y6	O8x4y6 escherichia
153	7	1.3	137	4	Q96AT1	Q96at1 homo sapien	226	7	1.3	229	16	O8X4V4	O8x4v4 escherichia
154	7	1.3	138	16	Q8CY70	Q8cy70 streptococc	227	7	1.3	229	16	O8X4P6	O8x4p6 escherichia
155	7	1.3	139	16	Q97ND0	Q97nd0 streptococc	228	7	1.3	229	16	O8X4R20	O8x4r20 shigella fl
156	7	1.3	140	5	O16685	O16685 caenorhabdi	229	7	1.3	229	16	O7UD0	O7ud0 shigella fl
157	7	1.3	141	15	O16685	O16685 caenorhabdi	230	7	1.3	231	16	O83R58	O83r58 shigella fl
158	7	1.3	142	15	O9JST2	O9jet2 human immun	231	7	1.3	232	16	O97D35	O97d35 clostridium
159	7	1.3	143	16	O8S6B2	O8s6b2 oryza sativ	232	7	1.3	232	17	O97Q73	O97q73 arabidopsis
160	7	1.3	144	17	O59371	O59371 pyrococcus	233	7	1.3	235	10	Q9LQ73	Q9lq73 arabidopsis
161	7	1.3	145	16	O7WM77	O7wm77 bordetella	234	7	1.3	236	4	Q7Z4I3	Q7z4i3 homo sapien
162	7	1.3	146	16	Q7W8L5	Q7w8l5 bordetella	235	7	1.3	237	16	Q8UBV8	Q8ubv8 agrobacteri


```
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Human skeletal muscle;
RA Koehler K., Beyer A., Mewes H.W., Weil B., Amid C., Osanger A.,
RA Fobo G., Han M., Wiemann S.;
RL Submitted (MAY-2003) to the EMBL/GenBank/DBJ databases.
DR ENBL: AL832338; CAD91138.1; -.
KW Hypothetical protein.
SQ SEQUENCE 546 AA; 61892 MW; 72867478997DF6 CRC64;

Query Match 77.7%; Score 412; DB 4; Length 546;
Best Local Similarity 99.8%; Pred. No. 0;
Matches 512; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 KSSPGQPEAGPEGAQERPSQAAPAVEAEGPGSSQAAPRKPGEQAQTAQSGALRDVSEELS 60
Db 17 KSSPGQPEAGPEGAQERPSQAAPAVEAEGPGSSQAAPRKPGEQAQTAQSGALRDVSEELS 76
QY 61 RQLEDILSTYCVNNQGGPGEQAQERPSQAAPAVEAEGPGSSQAAPRKPGEQAQTAQSGALRDVSEELS 120
Db 77 RQLEDILSTYCVNNQGGPGEQAQERPSQAAPAVEAEGPGSSQAAPRKPGEQAQTAQSGALRDVSEELS 136
QY 121 KGDPTNTEIRQSDVEGDRHRRPQSKKAKGLGKEITLLMOTLNTLSTPEEKLAALCKKY 180
Db 137 KGDPTNTEIRQSDVEGDRHRRPQSKKAKGLGKEITLLMOTLNTLSTPEEKLAALCKKY 196
QY 181 ABLLSEHNSQOMKLLQKQSOVLQOEKDLRGHSHKAVLARSKLSLCELOHNRSLK 240
Db 197 ABLLSEHNSQOMKLLQKQSOVLQOEKDLRGHSHKAVLARSKLSLCELOHNRSLK 256
QY 241 REGVQARAEERKKEVTSHFQVTLNDIQLQMSQHNRNSKLQENMELAEKLLIEQY 300
Db 257 REGVQARAEERKKEVTSHFQVTLNDIQLQMSQHNRNSKLQENMELAEKLLIEQY 316
QY 301 ELREEHIDKVPKHDLQOOLVDAKLQQAQEMLKEAERHOREKDFLLKEAVESQRMCELM 360
Db 317 ELREEHIDKVPKHDLQOOLVDAKLQQAQEMLKEAERHOREKDFLLKEAVESQRMCELM 376
QY 361 KQOETHLQKQALYATEKFEFQNTLSKSEVFTTFQEMEKMTKKIKLEKETTMYRSRW 420
Db 377 KQOETHLQKQALYATEKFEFQNTLSKSEVFTTFQEMEKMTKKIKLEKETTMYRSRW 436
QY 421 ESSNKALLEMAEKTVDKLEGLQVKIQRLEKLCRALQTERNDLNKRVQDLSAGGQSSL 480
Db 437 ESSNKALLEMAEKTVDKLEGLQVKIQRLEKLCRALQTERNDLNKRVQDLSAGGQSSL 496
QY 481 TDSGPERRPGGAQAPSSPRVTEAPCYPGAPS 513
Db 497 TDSGPERRPGGAQAPSSPRVTEAPCYPGAPS 529

RESULT 3
Q86T86 PRELIMINARY; PRT; 546 AA.
AC Q86T86;
DT 01-JUN-2003 (TrEMBLrel. 24, Created)
DT 01-JUN-2003 (TrEMBLrel. 24, Last sequence update)
DE Hypothetical protein DKFZp451I0918.
GN DKFZp451I0918.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RA Ansorge W., Krieger S., Mewes H.W., Weil B., Amid C., Fobo G., Han M.,
RA Osanger A., Wiemann S.;
RL Submitted (APR-2003) to the EMBL/GenBank/DBJ databases.
DR ENBL: AL832636; CAD89951.1; -.
KW Hypothetical protein.

Query Match 33.6%; Score 178; DB 4; Length 546;
Best Local Similarity 100.0%; Pred. No. 9.5e-175;
Matches 178; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 353 SQRMCELMKQOETHLQKQALYATEKFEFQNTLSKSEVFTTFQEMEKMTKKIKLEKE 412
Db 369 SQRMCELMKQOETHLQKQALYATEKFEFQNTLSKSEVFTTFQEMEKMTKKIKLEKE 428
```

QY 413 TTWYRSRWSSNKALLEMAEKTVDKLEGLQVKIQRLKLCALQTERNDLNKRVQDL 472
DB 429 TTWYRSRWSSNKALLEMAEKTVDKLEGLQVKIQRLKLCALQTERNDLNKRVQDL 488

QY 473 SAGGQSLTDSGPRRPGGAPQAPSSPRVTAPCYPGAPSTPTEASGOTGPQPTTSARA 530
DB 489 SAGGQSLTDSGPRRPGGAPQAPSSPRVTAPCYPGAPSTPTEASGOTGPQPTTSARA 546

RESULT 5
Q86Y86 PRELIMINARY; PRT; 116 AA.

ID Q86Y86
AC Q86Y86
DT 01-JUN-2003 (TREMELrel. 24, Created)
DT 01-JUN-2003 (TREMELrel. 24, Last sequence update)
DT 01-JUN-2003 (TREMELrel. 24, Last annotation update)
DE Hypothetical protein.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Uterus;
RA Strausberg R.;
RL Submitted (FEB-2003) to the EMBL/GenBank/DBJ databases.
DR EMBL; BC046565; AAH46565.1; -
KW Hypothetical protein.
SQ SEQUENCE 116 AA; 12623 MW; B99B79EACAA843F CRC64;

Query Match 21.9%; Score 116; DB 4; Length 116;
Best Local Similarity 100.0%; Pred. No. 3e-111;
Matches 116; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 415 MYRSRWSSNKALLEMAEKTVDKLEGLQVKIQRLKLCALQTERNDLNKRVQDL 474
DB 1 MYRSRWSSNKALLEMAEKTVDKLEGLQVKIQRLKLCALQTERNDLNKRVQDL 60

QY 475 GGQSLTDSGPRRPGGAPQAPSSPRVTAPCYPGAPSTPTEASGOTGPQPTTSARA 530
DB 61 GGQSLTDSGPRRPGGAPQAPSSPRVTAPCYPGAPSTPTEASGOTGPQPTTSARA 116

RESULT 6
Q8C5K1 PRELIMINARY; PRT; 150 AA.

ID Q8C5K1
AC Q8C5K1
DT 01-MAR-2003 (TREMELrel. 23, Created)
DT 01-MAR-2003 (TREMELrel. 23, Last sequence update)
DT 01-MAR-2003 (TREMELrel. 23, Last annotation update)
DE Hypothetical protein (Fragment).
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=Olfactory brain;
RX MEDLINE=22354683; PubMed=12466851;
RA The FANTOM Consortium,
RA the RIKEN Genome Exploration Research Group Phase I & II Team;
RT "Analysis of the mouse transcriptome based on functional annotation of
60,770 full-length cDNAs."
RL Nature 420:563-573 (2002).
DR EMBL; AK078192; BAC37168.1; -
KW Hypothetical protein.
FT NON TER 150
SQ SEQUENCE 150 AA; 16490 MW; C444717503E42E49 CRC64;

Query Match 4.0%; Score 21; DB 11; Length 150;
Best Local Similarity 100.0%; Pred. No. 6.8e-13;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 59 LSRQLEDILTTCYVDNNQGGP 79
DB 66 LSRQLEDILTTCYVDNNQGGP 86

RESULT 7
Q8BWJ7 PRELIMINARY; PRT; 241 AA.

ID Q8BWJ7
AC Q8BWJ7
DT 01-MAR-2003 (TREMELrel. 23, Created)
DT 01-MAR-2003 (TREMELrel. 23, Last sequence update)
DT 01-JUN-2003 (TREMELrel. 24, Last annotation update)
DE Muscle-derived protein MDP77 variant 1 (Fragment).
GN 2310001N14RIK.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=Heart;
RX MEDLINE=22354683; PubMed=12466851;
RA The FANTOM Consortium,
RA the RIKEN Genome Exploration Research Group Phase I & II Team;
RT "Analysis of the mouse transcriptome based on functional annotation of
60,770 full-length cDNAs."
RL Nature 420:563-573 (2002).
DR EMBL; AK052324; BAC34937.1; -
DR MGD; MGI:1916756; 2310001N14RIK.
FT NON TER 241
SQ SEQUENCE 241 AA; 27232 MW; 0B2DD1BD215630BB CRC64;

Query Match 3.6%; Score 19; DB 11; Length 241;
Best Local Similarity 100.0%; Pred. No. 1.2e-10;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 218 AVLARSKLESICRELQRHN 236
DB 210 AVLARSKLESICRELQRHN 228

RESULT 8
Q8BP11 PRELIMINARY; PRT; 241 AA.

ID Q8BP11
AC Q8BP11
DT 01-MAR-2003 (TREMELrel. 23, Created)
DT 01-MAR-2003 (TREMELrel. 23, Last sequence update)
DT 01-OCT-2003 (TREMELrel. 25, Last annotation update)
DE Hypothetical protein.
GN RBBP7.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=Mullerian duct includes surrounding region;
RX MEDLINE=22354683; PubMed=12466851;
RA The FANTOM Consortium,
RA the RIKEN Genome Exploration Research Group Phase I & II Team;
RT "Analysis of the mouse transcriptome based on functional annotation of
60,770 full-length cDNAs."
RL Nature 420:563-573 (2002).
DR EMBL; AK078477; BAC37296.1; -
DR MGD; MGI:1194910; Rbbp7.
DR GO; GO:0005634; C:nucleus; IDA.
DR GO; GO:0016564; F:transcriptional repressor activity; IDA.
DR GO; GO:0001222; P:negative regulation of transcription from P...; IDA.
KW Hypothetical protein.
SQ SEQUENCE 241 AA; 27598 MW; 1E0C7A38E169F31A CRC64;

Query Match 3.6%; Score 19; DB 11; Length 241;
Best Local Similarity 100.0%; Pred. No. 1.2e-10;

```
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 163 LNTLSTPEKLAALCKKYA 181
  |||||
  97 LNTLSTPEKLAALCKKYA 115

RESULT 9
Q8BUK2
ID Q8BUK2 PRELIMINARY; PRT; 463 AA.
AC Q8BUK2;
DT 01-MAR-2003 (TrEMBLrel. 23, Created)
DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE Muscle-derived protein MDP77 variant 1 (Fragment).
GN 231000IN14RIK.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
R2 SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=Heart;
RX MEDLINE=22354683; PubMed=12466851;
RA The FANTOM Consortium,
RA the RIKEN Genome Exploration Research Group Phase I & II Team;
RT "Analysis of the mouse transcriptome based on functional annotation of
RT 60,770 full-length cDNAs.";
RL Nature 420:563-573(2002).
DR EMBL; AK084639; BAC39238.1; --
DR MGD; MGI:1916756; 231000IN14RIK.
FT NON TER 463 463
SQ SEQUENCE 463 AA; 54090 MW; 1B8F8F81D73D58AA CRC64;

Query Match 3.6%; Score 19; DB 11; Length 463;
Best Local Similarity 100.0%; Pred. No. 2.2e-10;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 218 AVLARKSLSCRELQRHN 236
  |||||
  210 AVLARKSLSCRELQRHN 228

RESULT 10
Q8BHN1
ID Q8BHN1 PRELIMINARY; PRT; 524 AA.
AC Q8BHN1;
DT 01-MAR-2003 (TrEMBLrel. 23, Created)
DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Hypothetical protein.
GN RBP7.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
R2 SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=Cortex, Head, and Testis;
RX MEDLINE=22354683; PubMed=12466851;
RA The FANTOM Consortium,
RA the RIKEN Genome Exploration Research Group Phase I & II Team;
RT "Analysis of the mouse transcriptome based on functional annotation of
RT 60,770 full-length cDNAs.";
RL Nature 420:563-573(2002).
DR EMBL; AK030100; BAC26785.1; --
DR EMBL; AK031783; BAC27547.1; --
DR EMBL; AK044130; BAC31791.1; --
DR MGD; MGI:1194910; Rbp7.
DR GO; GO:0005634; C:nucleus; IDA.
DR GO; GO:0016564; F:transcriptional repressor activity; IDA.
DR GO; GO:0006122; P:negative regulation of transcription from p...; IDA.
KW Hypothetical protein.
```

```
SQ SEQUENCE 524 AA; 60308 MW; 0228777633E4ED7C CRC64;

Query Match 3.6%; Score 19; DB 11; Length 524;
Best Local Similarity 100.0%; Pred. No. 2.5e-10;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 163 LNTLSTPEKLAALCKKYA 181
  |||||
  146 LNTLSTPEKLAALCKKYA 164

RESULT 11
Q9NUQ3
ID Q9NUQ3 PRELIMINARY; PRT; 528 AA.
AC Q9NUQ3;
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT 01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
DE Hypothetical protein FLJ11209.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
R2 SEQUENCE FROM N.A.
RC TISSUE=Placenta;
RA Isogai T., Ota T., Hayaashi K., Sugiyama T., Otsuki T., Suzuki Y.,
RA Nishikawa T., Nagai K., Sugano S., Shiratori A., Sudo H.,
RA Watsutsuma M., Hosoi T., Kaku Y., Kodaira H., Kondo H., Sugawara M.,
RA Takahashi M., Chiba Y., Ishida S., Murakawa K., Ono Y., Takiguchi S.,
RA Yatanabe S., Kimura K., Murakami K., Ishii S., Kawai Y., Saito K.,
RA Yamamoto J., Wakamatsu A., Nakamura Y., Nagahari K., Masuho Y.,
RA Ninomiya K., Iwayanagi T.;
RT "NEDO human cDNA sequencing project.";
RL Submitted (FEB-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AK002071; BAA92068.1; --
DR Genew; HGNC:18578; CXorf15.
KW Hypothetical protein.
SQ SEQUENCE 528 AA; 60605 MW; CA88D781DE06ACB3 CRC64;

Query Match 3.6%; Score 19; DB 4; Length 528;
Best Local Similarity 100.0%; Pred. No. 2.5e-10;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 163 LNTLSTPEKLAALCKKYA 181
  |||||
  146 LNTLSTPEKLAALCKKYA 164

RESULT 12
Q8VET1
ID Q8VET1 PRELIMINARY; PRT; 584 AA.
AC Q8VET1;
DT 01-MAR-2002 (TrEMBLrel. 20, Created)
DT 01-MAR-2002 (TrEMBLrel. 20, Last sequence update)
DT 01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
DE Muscle-derived protein MDP77 variant 2 (Muscle-derived protein MDP77
DE variant 1).
GN 231000IN14RIK OR MDP77.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
R2 SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=Heart;
RX MEDLINE=21653722; PubMed=11805063;
RA Benson K.F., Chada K.;
RT "Molecular Characterization of the Mouse In(10)17Rk Inversion and
RT Identification of a Novel Muscle-Specific Gene at the Proximal
RT Breakpoint.";
RL Genetics 160:279-287(2002).
DR EMBL; AF422245; AAL33910.1; --
```


DR EMBL; AF422244; AAL33909.1; -.
DR MGP; MGI:1916756; 2310001N14RIK.
SQ SEQUENCE 684 AA; 77049 MW; E8A64E28EF56E31 CRC64;

Query Match 3.6%; Score 19; DB 11; Length 684;
Best Local Similarity 100.0%; Pred. No. 3.1e-10;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 218 AVLARSKLESCLRELQRHN 236
DB 209 AVLARSKLESCLRELQRHN 227
|||||

RESULT 13
Q9POX1 PRELIMINARY; PRT; 186 AA.
AC Q9POX1, 20, Last sequence update)
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-MAR-2002 (TrEMBLrel. 20, Last sequence update)
DT 01-MAR-2002 (TrEMBLrel. 20, Last sequence update)
DE Lipopolysaccharide specific response-5 protein (Fragment).
GN LSR5.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RA Chai Y.B., Zhao Z.L., Zhu P., Yan W., Chen N.C., Wang Q., Yue L.,
Chen S.M.; sapiens gene from dental pulp cells."
RT "New Homo sapiens gene from dental pulp cells."
RL Submitted (DEC-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF143740; AAF70546.2; -.
FT NON TER 1
SQ SEQUENCE 186 AA; 21464 MW; 93BF8383096B0A CRC64;

Query Match 3.2%; Score 17; DB 4; Length 186;
Best Local Similarity 100.0%; Pred. No. 1.1e-08;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 397 QEMKMTKKIKLEKET 413
DB 38 QEMKMTKKIKLEKET 54
|||||

RESULT 14
Q8N3S2 PRELIMINARY; PRT; 505 AA.
AC Q8N3S2, 22, Created)
DT 01-OCT-2002 (TrEMBLrel. 22, Last sequence update)
DT 01-OCT-2002 (TrEMBLrel. 22, Last sequence update)
DT 01-OCT-2002 (TrEMBLrel. 22, Last sequence update)
DE Hypothetical protein (Fragment).
GN DKFPZ451A175.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RA Koehrer X., Beyer A., Mewes H.W., Weil B., Wiemann S.;
Submitted (JUL-2002) to the EMBL/GenBank/DBJ databases.
RL EMBL; AL832322; CAD38617.1; -.
DR EMBL; AL832322; CAD38617.1; -.
KW Hypothetical protein.
FT NON TER 1
SQ SEQUENCE 505 AA; 56957 MW; 1AA33548A2F1DADE CRC64;

Query Match 3.2%; Score 17; DB 4; Length 505;
Best Local Similarity 100.0%; Pred. No. 2.8e-08;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 220 LARSKLESCLRELQRHN 236
|||||

Db 30 LARSKLESCLRELQRHN 46

RESULT 15
Q91969 PRELIMINARY; PRT; 676 AA.
AC Q91969, 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE Muscle derived protein.
GN MDP77.
OS Gallus gallus (Chicken).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Archosauria; Aves; Neognathae; Galliformes; Phasianinae;
OC Gallus.
OX NCBI_TaxID=9031;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Muscle;
RA MBLIN=20175243; PubMed=10708594;
RA Uyeda A., Fukui I., Fujimori K., Kiyosue K., Kishimune H., Kasai M.,
Taguchi T.;
RT "MDP77: A novel neurite-outgrowth-promoting protein predominantly
expressed in chick muscles."
RL Biochem. Biophys. Res. Commun. 269:564-569(2000).
DR EMBL; D89399; BAA94755.1; -.
DR PIR; JC7222; JC7222.
SQ SEQUENCE 676 AA; 77020 MW; FCEA9E393250EE94 CRC64;

Query Match 3.2%; Score 17; DB 13; Length 676;
Best Local Similarity 100.0%; Pred. No. 3.6e-08;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 220 LARSKLESCLRELQRHN 236
DB 207 LARSKLESCLRELQRHN 223
|||||

RESULT 16
Q8N3L3 PRELIMINARY; PRT; 715 AA.
AC Q8N3L3, 22, Created)
DT 01-OCT-2002 (TrEMBLrel. 22, Last sequence update)
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE Hypothetical protein DKFPZ451F022 (Fragment).
GN DKFPZ451F022.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RA Wambutt R., Heubner D., Mewes H.W., Weil B., Amid C., Fobo G., Han M.,
Csanger A., Wiemann S.;
Submitted (APR-2003) to the EMBL/GenBank/DBJ databases.
RL EMBL; AL834248; CAD38924.2; -.
DR EMBL; AL834248; CAD38924.2; -.
KW Hypothetical protein.
FT NON TER 1
SQ SEQUENCE 715 AA; 79798 MW; 11C50191BCD26582 CRC64;

Query Match 3.2%; Score 17; DB 4; Length 715;
Best Local Similarity 100.0%; Pred. No. 3.8e-08;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 220 LARSKLESCLRELQRHN 236
DB 240 LARSKLESCLRELQRHN 256
|||||

RESULT 17
Q86T52

ID Q86T52 PRELIMINARY; PRT; 718 AA.
AC Q86T52;
DT 01-JUN-2003 (TrEMBLrel. 24, Created)
DT 01-JUN-2003 (TrEMBLrel. 24, Last sequence update)
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE Hypothetical protein DKFZp451G083 (Fragment).
GN DKFZP451G083.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Human skeletal muscle;
RA Wambutt R., Heubner D., Mewes H.W., Weil B., Amid C., Osanger A.,
RA Fobo G., Han M., Wiemann S.;
RL Submitted (MAY-2003) to the EMBL/GenBank/DBJ databases.
DR EMBL; AL811988; CAD91140.1; -;
KW Hypothetical protein.
FT NON TER 1
SQ SEQUENCE 718 AA; 80290 MW; 16C3BP22C19559A4 CRC64;
Query Match 3.2%; Score 17; DB 4; Length 718;
Best Local Similarity 100.0%; Pred. No. 3.8e-08;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 220 LARSKLSLCRELQHN 236
DB 243 LARSKLSLCRELQHN 259
RESULT 18
Q8RXD7 PRELIMINARY; PRT; 404 AA.
ID Q8RXD7;
AC Q8RXD7;
DT 01-JUN-2002 (TrEMBLrel. 21, Created)
DT 01-JUN-2002 (TrEMBLrel. 21, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Hypothetical protein (At5G50840).
GN AT5G50840.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids;
OC eurosids II; Brassicales; Brassicaceae; Arabidopsis.
OX NCBI_TaxID=3702;
RN [1]
RP SEQUENCE FROM N.A.
RA Nguyen M., Karlin-Neumann G., Southwick A., Lam B., Miranda M.,
RA Palm C.J., Bowser L., Jones T., Banh J., Carninci P., Chen H.,
RA Cheuk R., Chung M.K., Hayashizaki Y., Ishida J., Kamiya A., Kawai J.,
RA Kim C., Lin J., Liu S.X., Narusaka M., Pham P.K., Sakano H.,
RA Sakurai T., Satou M., Seki M., Shinn P., Yamada K., Shinozaki K.,
RA Ecker J., Theologis A., Davis R.W.;
RL Submitted (FEB-2002) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RA Cheuk R., Chen H., Kim C.J., Shinn P., Bowser L., Carninci P.,
RA Dale J.M., Hayashizaki Y., Ishida J., Jones T., Kamiya A.,
RA Karlin-Neumann G., Kawai J., Lam B., Lin J., Miranda M., Narusaka M.,
RA Nguyen M., Onodera C.S., Palm C.J., Quach H.L., Sakurai T., Satou M.,
RA Seki M., Southwick A., Toriumi M., Wong C., Wu H.C., Yamada K., Yu G.,
RA Shinozaki K., Davis R.W., Theologis A., Ecker J.R.;
RT "Arabidopsis ORF clones";
RL Submitted (JUN-2003) to the EMBL/GenBank/DBJ databases.
DR EMBL; AY081318; RAL01207.1; -;
DR EMBL; BT008857; AAP68296.1; -;
KW Hypothetical protein.
SQ SEQUENCE 404 AA; 46796 MW; 3C31C38A7E437DF6 CRC64;
Query Match 2.1%; Score 11; DB 10; Length 404;
Best Local Similarity 100.0%; Pred. No. 0.036;
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 224 KLESCLRELQ 234
DB 163 KLESCLRELQ 173
RESULT 19
Q8BZA4 PRELIMINARY; PRT; 181 AA.
ID Q8BZA4;
AC Q8BZA4;
DT 01-JUN-2001 (TrEMBLrel. 17, Created)
DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE Lipopolysaccharide-specific response 5-like protein (Hypothetical
protein DKFZp451G0616).
GN DKFZP451G0616.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RA Kuroda-Kawaguchi T., Skaletsky H., Minx P.J., Brown L.G., Rozen S.,
RA Wilson R.K., Waterston R.H., Page D.C.;
RT "The DNA sequence of the human Y chromosome";
RL Submitted (DEC-2000) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RA Ansorge W., Krieger S., Mewes H.W., Weil B., Amid C., Fobo G., Han M.,
RA Osanger A., Wiemann S.;
RL Submitted (APR-2003) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF332225; RAK13477.1; -;
DR EMBL; AL832583; CAD89940.1; -;
DR Genew; HGNC:18577; Cyorf159.
KW Hypothetical protein.
SQ SEQUENCE 181 AA; 20858 MW; 4BA8B930933012A1 CRC64;
Query Match 1.9%; Score 10; DB 4; Length 181;
Best Local Similarity 100.0%; Pred. No. 0.19;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 403 TKKIKKLEKE 412
DB 44 TKKIKKLEKE 53
RESULT 20
Q8BUK8 PRELIMINARY; PRT; 190 AA.
ID Q8BUK8;
AC Q8BUK8;
DT 01-MAR-2003 (TrEMBLrel. 23, Created)
DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE Muscle-derived protein MDP77 variant 1.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=Heart;
RX MEDLINE=22354683; PubMed=12466851;
RA The FANTOM Consortium,
RA the RIKEN Genome Exploration Research Group Phase I & II Team;
RT "Analysis of the mouse transcriptome based on functional annotation of
60,770 full-length cDNAs";
RL Nature 420:563-573(2002).
DR EMBL; AK084579; BAC39219.1; -;
SQ SEQUENCE 190 AA; 20852 MW; D96330563C9830D6 CRC64;
Query Match 1.9%; Score 10; DB 11; Length 190;
Best Local Similarity 100.0%; Pred. No. 0.2;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 57 EELSRLQLEDI 66
|||||
Db 60 EELSRLQLEDI 69

RESULT 21

Q8DTC1 PRELIMINARY; PRT; 637 AA.
AC Q8DTC1;
DT 01-MAR-2003 (TrEMBLrel. 23, Created)
DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Putative ABC transporter, ATP-binding protein.
GN SWU.1431C.
OS Streptococcus mutans.
OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
OC Streptococcus.
OX NCBI_TaxId=1309;
RN [1]
SEQUENCE FROM N.A.
RC STRAIN=UAI59 / ATCC 700610 / Serotype C;
RX MEDLINE=22295063; PubMed=12397186;
RA Ajdic D., McShan W.M., McLaughlin R.E., Savic G., Chang J.,
RA Carson M.B., Primeaux C., Tian R., Kenton S., Jia H., Lin S., Qian Y.,
RA Li S., Zhu H., Najjar F., Lai H., White J., Roe B.A., Ferretti J.J.;
RT "Genome sequence of Streptococcus mutans UAI59, a cariogenic dental
pathogen.";
RL Proc. Natl. Acad. Sci. U.S.A. 99:14434-14439(2002).
DR EMBL; AE014976; AAM59095.1; IEA.
DR GO; GO:0016020; C:membrane; IEA.
DR GO; GO:0005524; P:ATP binding; IEA.
DR GO; GO:0004009; P:ATP-binding cassette (ABC) transporter acti. . . ; IEA.
DR GO; GO:0000166; P:nucleotide binding; IEA.
DR GO; GO:0006810; P:transport; IEA.
DR InterPro; IPR003593; AAA_ATPase.
DR Pfam; PF00005; ABC_tran; 2.
DR ProDom; PD000006; ABC_transporter; 2.
DR SMART; SM00382; AAA; 2.
DR PROSITE; PS00211; ABC_TRANSPORTER_1; 1.
DR PROSITE; PS00893; ABC_TRANSPORTER_2; 2.
KW ATP-binding; Complete proteome.
SQ SEQUENCE 637 AA; 72234 MW; CB75C5FB8716B37 CRC64;

Query Match 1.7%; Score 9; DB 16; Length 637;
Best Local Similarity 100.0%; Pred. No. 6.4;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 57 EELSRLQLED 65
|||||
Db 629 EELSRLQLED 637

RESULT 22

Q9VPS3 PRELIMINARY; PRT; 826 AA.
AC Q9VPS3;
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-OCT-2002 (TrEMBLrel. 22, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE CG2839 protein.
GN CG2839.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxId=7227;
RN [1]
SEQUENCE FROM N.A.
RC STRAIN=Berkely;
RX MEDLINE=20196006; PubMed=107311132;
RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA Amanatides P.G., Scher S.E., Li P.W., Hoskins R.A., Galle R.F.,

George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,
Wan K.H., Doyle C., Baxter B.G., Helt G., Nelson C.R., Miklos G.L.G.,
April J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
Ballaw K.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
Beeson K.Y., Benos P.V., Berman J.P., Bhandari D., Bolshakov S.,
Borkova D., Botchan M.R., Bouck J., Brockstein P., Brotter P.,
Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
de la Cruz B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,
Foster C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
Glodek A., Gong F., Gottrell J.H., Gu Z., Guan P., Harris M.,
Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
Hoskins D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
Kimmel B.E., Klotzel C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacle J.M.,
Palazzo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
Reichert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
Svirskas R., Tector C., Turner R., Venter G., Wang A.H., Wang X.,
Wang Z.-Y., Wasserman D.A., Weinstock G.M., Weissbach J.,
Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao O.A.,
Ye J., Yeh R.F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
Zheng X.H., Zhong F.N., Zhou W., Zhou S., Zhu S., Zhu X., Smith H.O.,
Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT "The genome sequence of Drosophila melanogaster.";
RL Science 287:2185-2195(2000).
RN [2]
RP SEQUENCE FROM N.A.
RA Celniker S.E., Adams M.D., Krommiller B., Wan K.H., Holt R.A.,
RA Evans C.A., Gocayne J.D., Amanatides P.G., Brandon R.C., Rogers Y.,
RA Banzon J., An H., Baldwin D., Banzon J., Beeson K.Y., Sussam D.A.,
RA Carlson J.W., Center A., Champe M., Davenport L.B., Dietz S.M.,
RA Dodson K., Dorsett V., Doup L.E., Doyle C., Dresnek D., Farfan D.,
RA Ferreira S., Frise E., Galle R.F., Garg N.S., George R.A.,
RA Gonzalez M., Houck J., Hoskins R.A., Hostin D., Howland T.J.,
RA Ibegwam C., Jalali M., Kruse D., Li P., Mattei B., Moshrefi A.,
RA McIntosh T.C., Moy M., Murphy B., Nelson C., Nelson K.A., Nunco J.,
RA Pacle J., Paragas V., Park S., Patel S., Pfeiffer B.,
RA Phouanavong S., Pittman G.S., Puri V., Richards S., Scheeler F.,
RA Stapleton M., Strong R., Svirskas R., Tector C., Tyler D.,
RA Williams S.M., Zaveri J.S., Smith H.O., Venter J.C., Rubin G.M.;
RT "Sequencing of Drosophila melanogaster genome.";
RL Submitted (MAR-2000) to the EMBL/GenBank/DBSJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RA Misra S., Crosby M.A., Matthews B.B., Bayraktaroglu L., Campbell K.,
RA Hradecky P., Huang Y., Kaminker J.S., Prochuk S.E., Smith C.D.,
RA Tupy J.L., Bergman C., Bernier B., Carlson J.W., Celniker S.E.,
RA Clamp M., Drysdale R., Emmert D., Frise E., de Grey A., Harris N.,
RA Krommiller B., Marshall B., Millburn G., Richter J., Russo S.,
RA Searle S.M.J., Smith E., Shu S., Smutnick F., Whitfield S.,
RA Ashburner M., Gelbart W.M., Rubin G.M., Mungall C.J., Lewis S.E.;
RT "Annotation of Drosophila melanogaster genome.";
RL Submitted (MAR-2000) to the EMBL/GenBank/DBSJ databases.
RN [4]
RP SEQUENCE FROM N.A.
RA Adams M.D., Celniker S.E., Gibbs R.A., Rubin G.M., Venter C.J.;
RL Submitted (MAR-2000) to the EMBL/GenBank/DBSJ databases.
RN [5]
RP SEQUENCE FROM N.A.
RA FlyBase;
RL Submitted (SEP-2002) to the EMBL/GenBank/DBSJ databases.
RL EMBL; AE003588; AAF51469.2; -
DR

DR HSP; P22897; IBGG.
 DR FlyBase; FBgn0031273; CG2839.
 DR GO; GO:0005529; F:sugar binding; IEA.
 DR InterPro; IPR001304; Lectin_C.
 DR Pfam; PF00059; lectin_c; 1.
 DR SMART; SM00034; CLCT_1.
 DR PROSITE; PS00041; C-TYPE LECTIN 2; 1.
 SQ SEQUENCE 826 AA; 106026 MW; B2C8A017DB42716C CRC64;

Query Match 1.7%; Score 9; DB 5; Length 826;
 Best Local Similarity 100.0%; Pred. No. 8.1;
 Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 248 REEEERKR 256
 Db 488 REEEERKR 496
 |||||

RESULT 23
 Q96WP6 PRELIMINARY; PRT; 842 AA.
 AC Q96WP6
 DT 01-DEC-2001 (TrEMBLrel. 19, Created)
 DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
 DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
 DE CDC5.
 GN CDC5.
 OS Lentinula edodes (Shiitake mushroom) (Lentinus edodes).
 OC Eukaryota; Fungi; Basidiomycota; Hymenomycetes; Homobasidiomycetes;
 OC Agaricales; Tricholomataceae; Lentinula.
 OX NCBI_TaxID=5353;
 [1]
 RP SEQUENCE FROM N.A.
 RA Miyazaki Y., Shishido K.;
 RT "cDNA sequence of L.edodes Le.cdc5";
 RL Submitted (AUG-2001) to the EMBL/GenBank/DBJ databases.
 CC -1- SUBCELLULAR LOCATION: NUCLEAR (BY SIMILARITY).
 CC -1- SIMILARITY: CONTAINS 2 MYB-LIKE DOMAINS.
 DR EMBL; AB067691; BAB62527.1; -;
 DR GO; GO:0005634; C:nucleus; IEA.
 DR GO; GO:0003677; F:DNA binding; IEA.
 DR InterPro; IPR001005; Myb_DNA_binding.
 DR Pfam; PF00249; myb_DNA_binding; 2.
 DR SMART; SM00717; SANT; 2.
 DR PROSITE; PS00037; MYB_1; 1.
 DR PROSITE; PS00334; MYB_2; 1.
 DR PROSITE; PS00090; MYB_3; 2.
 DR DNA-binding; Nuclear Protein.
 KW DNA-binding; Nuclear Protein.
 SQ SEQUENCE 842 AA; 93740 MW; 273B3592B5BAE589 CRC64;

Query Match 1.7%; Score 9; DB 3; Length 842;
 Best Local Similarity 100.0%; Pred. No. 8.3;
 Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 246 REEEERKR 254
 Db 538 REEEERKR 546
 |||||

RESULT 24
 Q861Z2 PRELIMINARY; PRT; 113 AA.
 AC Q861Z2
 DT 01-JUN-2003 (TrEMBLrel. 24, Created)
 DT 01-JUN-2003 (TrEMBLrel. 24, Last sequence update)
 DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
 DE Hypothetical protein.
 OS Dictyostelium discoideum (Slime mold).
 OC Eukaryota; Mycetozoa; Dictyostellida; Dictyostelium.
 OX NCBI_TaxID=44689;
 [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=AX4;

RX MEDLINE=22092622; PubMed=12097910;
 RA Gloeckner G., Eichinger L., Szafranski K., Pachebat J., Dear P.,
 RA Lehmann R., Baumgart C., Parra G., April J.F., Guigo R., Kumpf K.,
 RA Tunggal B., Cox E., Quail M.A., Platzer M., Rosenthal A., Noegel A.A.;
 RT "Sequence and analysis of chromosome 2 of Dictyostelium discoideum";
 RL Nature 418:79-85(2002).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=AX4;
 RA Baumgart C.;
 RL Submitted (MAR-2003) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AC116986; AAO51852.1; -;
 KW Hypothetical protein.
 SQ SEQUENCE 113 AA; 12894 MW; 996243B5C6A38DA1 CRC64;

Query Match 1.5%; Score 8; DB 5; Length 113;
 Best Local Similarity 100.0%; Pred. No. 15;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 247 AREEEERKR 254
 Db 60 AREEEERKR 67
 |||||

RESULT 25
 Q8PL61 PRELIMINARY; PRT; 135 AA.
 AC Q8PL61
 DT 01-OCT-2002 (TrEMBLrel. 22, Created)
 DT 01-OCT-2002 (TrEMBLrel. 22, Last sequence update)
 DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
 DE Flagellar protein.
 GN FLIO OR XAC1945.
 OS Xanthomonas axonopodis (pv. citri).
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Xanthomonadales;
 OC Xanthomonadaceae; Xanthomonas.
 OX NCBI_TaxID=92829;
 [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=306 / ATCC 13902 / XV 101;
 RX MEDLINE=22022145; PubMed=12024217;
 RA da Silva A.C.R., Ferro J.A., Reinach P.C., Farah C.S., Furlan L.R.,
 RA Quaggio R.B., Monteiro-Vitorello C.B., Van Sluys M.A., Almeida N.F.,
 RA Alves L.M.C., do Amaral A.M., Bertolini M.C., Camargo L.E.A.,
 RA Camarotte G., Cannavan F., Cardoso J., Chamargo F., Ciapina L.P.,
 RA Cicarelli R.M.B., Coutinho L.L., Cursino-Santos J.R., El-Dorry H.,
 RA Faria J.B., Ferreira A.J.S., Ferreira R.C.C., Ferro M.I.T.,
 RA Formighieri E.F., Franco M.C., Greggio C.C., Gruber A.,
 RA Katsuyama A.M., Kishi L.T., Leite R.P., Lemos E.G.M., Lemos M.V.P.,
 RA Locali E.C., Machado M.A., Madeira A.M.B.N., Martinez-Rossi N.M.,
 RA Martins E.C., Meidanis J., Menck C.F.M., Miyaki C.Y., Moon D.H.,
 RA Moreira L.M., Novo M.T.M., Okura V.K., Oliveira M.C., Oliveira V.R.,
 RA Pereira H.A., Rossi A., Sena J.A.D., Silva C., de Souza R.F.,
 RA Spinola L.A.F., Takita M.A., Tamura R.B., Teixeira E.C., Tezza R.I.D.,
 RA Trindade dos Santos M., Truffi D., Tsai S.M., White P.F.,
 RA Setubal J.C., Kitajima J.P.;
 RT "Comparison of the genomes of two Xanthomonas pathogens with differing
 RT host specificities";
 RL Nature 417:459-463(2002).
 DR EMBL; AB011830; AAM36807.1; -;
 DR InterPro; IPR007442; FLIO.
 DR Pfam; PF04347; FLIO; 1.
 KW Complete proteome.
 SQ SEQUENCE 135 AA; 13836 MW; 38A5CFB543045981 CRC64;

Query Match 1.5%; Score 8; DB 16; Length 135;
 Best Local Similarity 100.0%; Pred. No. 17;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 493 GAQAPSSP 500
 Db 19 GAQAPSSP 26
 |||||

DR EMBL; D88189; BAA13557.1; -.
 DR GO; GO:0019089; P:Viral transmission; IEA.
 DR InterPro; IPR007683; VapD_N.
 DR Pfam; PF04605; VapD_N; 1.
 DR SEQUENCE 143 AA; 16728 MW; 39C76D7F45834327 CRC64;
 SQ
 Query Match 1.5%; Score 8; DB 2; Length 143;
 Best Local Similarity 100.0%; Pred. No. 18;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 242 EGVQVARE 249
 |||||
 DB 87 EGVQVARE 94
 |||||

RESULT 28
 017194 PRELIMINARY; PRT; 195 AA.
 ID 017194
 AC 017194
 DT 01-JAN-1998 (TrEMBLrel. 05, Created)
 DT 01-OCT-2003 (TrEMBLrel. 23, Last sequence update)
 DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
 DE Hypothetical protein C08E3.4.
 GN C08E3.4.
 OS Caenorhabditis elegans.
 OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
 OC Rhabditidae; Telodermata; Caenorhabditis.
 OX NCBI_TaxID=6239;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=Bristol N2;
 RX MEDLINE=99069613; PubMed=9851916;
 RA Waterston R.;
 RT "Genome sequence of the nematode C. elegans: a platform for
 RT investigating biology. The C. elegans Sequencing Consortium.;
 RL Science 282:2012-2018 (1998).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=Bristol N2;
 RA Miller N., Kramer J., Keppeler D.;
 RT "The sequence of C. elegans cosmid C08E3.4";
 RL Submitted (SEP-1997) to the EMBL/GenBank/DBJ databases.
 RN [3]
 RP SEQUENCE FROM N.A.
 RC STRAIN=Bristol N2;
 RA Waterston R.;
 RL Submitted (NOV-2002) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AF025457; AAB70965.2; -.
 DR PIR; T32354; T32354.
 DR WormPep; C08E3.4; C832119.
 DR InterPro; IPR002900; DUF38.
 DR Pfam; PF01827; FTH; 1.
 DR Hypothetical protein.
 KW Hypothetical protein.
 SQ SEQUENCE 195 AA; 22534 MW; A750F3C8AD1BFA5F CRC64;
 Query Match 1.5%; Score 8; DB 5; Length 195;
 Best Local Similarity 100.0%; Pred. No. 24;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 290 ARLKKLI 297
 |||||
 DB 65 ARLKKLI 72
 |||||

RESULT 29
 083C21 PRELIMINARY; PRT; 206 AA.
 ID 083C21
 AC 083C21
 DT 01-JUN-2003 (TrEMBLrel. 24, Created)
 DT 01-JUN-2003 (TrEMBLrel. 24, Last sequence update)
 DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
 DE Hypothetical protein.
 GN CBU1314.

QY 493 GAQAPSP 500
 |||||
 DB 19 GAQAPSP 26
 |||||

RESULT 27
 P70717 PRELIMINARY; PRT; 143 AA.
 ID P70717
 AC P70717
 DT 01-FEB-1997 (TrEMBLrel. 02, Created)
 DT 01-FEB-1997 (TrEMBLrel. 02, Last sequence update)
 DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
 DE VapD-homolog.
 OS Actinobacillus actinomycetemcomitans (Haemophilus
 OS actinomycetemcomitans).
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Pasteurellales;
 OC Pasteurellaceae; Actinobacillus.
 OX NCBI_TaxID=714;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=Y4;
 RA MEDLINE=97148607; PubMed=9020051;
 RA Yoshida Y., Nakano Y., Yamashita Y., Koga T.;
 RT "The gnd gene encoding a novel 6-phosphogluconate dehydrogenase and
 RT its adjacent region of Actinobacillus actinomycetemcomitans
 RT chromosomal DNA."; Res. Commun. 230:220-225 (1997).
 RL Biochem. Biophys. Res. Commun. 230:220-225 (1997).

Query Match 1.5%; Score 8; DB 16; Length 135;
 Best Local Similarity 100.0%; Pred. No. 17;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 493 GAQAPSP 500
 |||||
 DB 19 GAQAPSP 26
 |||||

RESULT 28
 017194 PRELIMINARY; PRT; 195 AA.
 ID 017194
 AC 017194
 DT 01-JAN-1998 (TrEMBLrel. 05, Created)
 DT 01-OCT-2003 (TrEMBLrel. 23, Last sequence update)
 DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
 DE Hypothetical protein C08E3.4.
 GN C08E3.4.
 OS Caenorhabditis elegans.
 OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
 OC Rhabditidae; Telodermata; Caenorhabditis.
 OX NCBI_TaxID=6239;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=Bristol N2;
 RX MEDLINE=99069613; PubMed=9851916;
 RA Waterston R.;
 RT "Genome sequence of the nematode C. elegans: a platform for
 RT investigating biology. The C. elegans Sequencing Consortium.;
 RL Science 282:2012-2018 (1998).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=Bristol N2;
 RA Miller N., Kramer J., Keppeler D.;
 RT "The sequence of C. elegans cosmid C08E3.4";
 RL Submitted (SEP-1997) to the EMBL/GenBank/DBJ databases.
 RN [3]
 RP SEQUENCE FROM N.A.
 RC STRAIN=Bristol N2;
 RA Waterston R.;
 RL Submitted (NOV-2002) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AF025457; AAB70965.2; -.
 DR PIR; T32354; T32354.
 DR WormPep; C08E3.4; C832119.
 DR InterPro; IPR002900; DUF38.
 DR Pfam; PF01827; FTH; 1.
 DR Hypothetical protein.
 KW Hypothetical protein.
 SQ SEQUENCE 195 AA; 22534 MW; A750F3C8AD1BFA5F CRC64;
 Query Match 1.5%; Score 8; DB 16; Length 135;
 Best Local Similarity 100.0%; Pred. No. 17;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 493 GAQAPSP 500
 |||||
 DB 19 GAQAPSP 26
 |||||

RESULT 29
 083C21 PRELIMINARY; PRT; 206 AA.
 ID 083C21
 AC 083C21
 DT 01-JUN-2003 (TrEMBLrel. 24, Created)
 DT 01-JUN-2003 (TrEMBLrel. 24, Last sequence update)
 DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
 DE Hypothetical protein.
 GN CBU1314.

OS Cooxiella burnetii.
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Legionellales;
 OC Cooxiellaceae; Cooxiella.
 OX NCBI_TaxID=777;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=Nine Mile phase I / RSA 493;
 RX MEDLINE=22608657; PubMed=12704232;
 RA Seshadri R., Paulsen I.T., Eisen J.A., Read T.D., Nelson K.E.,
 RA Nelson W.C., Ward N.L., Tettelin H., Davidsen T.M., Beanan M.J.,
 RA DeBoy R.T., Daugherty S.C., Brinkac L.M., Madupu R., Dodson R.J.,
 RA Khouri H.M., Lee K.H., Carty H.A., Scanlan D., Heinzen R.A.,
 RA Thompson H.A., Samuel J.E., Fraser C.M., Heidelberg J.F.;
 RT "Complete genome sequence of the Q-fever pathogen, Cooxiella
 burnetii.";
 RL Proc. Natl. Acad. Sci. U.S.A. 100:5455-5460(2003).
 DR EMBL; AB016964; AAC090819.1; -;
 DR TIGR; CHU1314; -;
 RX HYPOTHEICAL protein; Complete proteome.
 XW SEQUENCE 206 AA; 23477 MW; 1C83C21054AED92D CRC64;
 SQ
 Query Match 1.5%; Score 8; DB 16; Length 206;
 Best Local Similarity 100.0%; Pred. No. 25;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 Qy 404 KXIKKLEK 411
 Db 69 KXIKKLEK 76
 RESULT 30
 Q9SCX3 PRELIMINARY; PRT; 224 AA.
 AC Q9SCX3;
 DT 01-MAY-2000 (TrEMBLrel. 13, Created)
 DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
 DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
 DE Elongation factor 1B alpha-subunit.
 GN EF1BAPHAZ OR T20D1_30/AT5G19510.
 OS Arabidopsis thaliana (Mouse-ear cress).
 OC Eukaryota; Viridiplantae; Streptophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids;
 OC eumids II; Brassicales; Brassicaceae; Arabidopsi.
 OX NCBI_TaxID=3702;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC MEDLINE=20686877; PubMed=10618495;
 RA Hexicourt F., Jupin I.;
 RT "Molecular cloning and characterization of the Arabidopsis thaliana
 alpha-subunit of elongation factor 1B.";
 RL FEBS Lett. 464:148-152(1999).
 RN [2]
 RP SEQUENCE FROM N.A.
 RA Yamada K., Liu S.X., Sakano H., Pham P.K., Banh J., Chung M.K.,
 RA Goldsmith A.D., Lee J.M., Quach H.L., Toriumi M., Yamamura Y., Yu G.,
 RA Carninci P., Chen H., Cheuk R., Hayashizaki Y., Ishida J., Jones T.,
 RA Kamiya A., Karlin-Neumann G., Kawai J., Kim C., Lam B., Lin J.,
 RA Miranda M., Narusaka M., Nguyen M., Palm C.J., Sakurai T., Satou M.,
 RA Seki M., Shinn P., Southwick A., Shinozaki K., Davis R.W., Ecker J.R.,
 RA Theologis A.;
 RT "Full length cDNA of gene T20D1_30/AT5G19510.";
 RL Submitted (MAR-2001) to the EMBL/GenBank/DBJ databases.
 RN [3]
 RP SEQUENCE FROM N.A.
 RA Yamada K., Banh J., Banno P., Dale J.M., Goldsmith A.D., Lee J.M.,
 RA Onodera C.S., Quach H.L., Tang C., Toriumi M., Yamamura Y., Yu G.,
 RA Yu S., Bowser L., Carninci P., Chen H., Cheuk R., Hayashizaki Y.,
 RA Ishida J., Jones T., Kamiya A., Karlin-Neumann G., Kawai J., Kim C.,
 RA Koesema E., Lam B., Lin J., Meyers M.C., Miranda M., Narusaka M.,
 RA Nguyen M., Palm C.J., Sakurai T., Satou M., Seki M., Shinn P.,
 RA Southwick A., Tracy S.E., Shinozaki K., Davis R.W., Ecker J.R.,
 RA Theologis A.;
 RT "Full length cDNA of gene T20D1_30/AT5G19510.";

RL Submitted (SEP-2001) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AJ249597; CAB64730.1; -;
 DR EMBL; AF360304; AAK26014.1; -;
 DR EMBL; AY056354; AAL07240.1; -;
 DR PIR; PA0110; PA0110.
 DR PIR; T52558; T52558.
 DR GO; GO:0005853; C:eukaryotic translation elongation factor 1.; IEA.
 DR GO; GO:0003746; P:translation elongation factor activity; IEA.
 DR GO; GO:0006414; P:translational elongation; IEA.
 DR InterPro; IPR001326; EFl_BD.
 DR InterPro; IPR004046; GST_Cterm.
 DR Pfam; PF00736; EFlBD; 1; -;
 DR Pfam; PF00043; GST_C; 1;
 DR PROSITE; PS00824; EFlBD_1; 1;
 DR PROSITE; PS00825; EFlBD_2; 1;
 KW Elongation factor.
 SQ SEQUENCE 224 AA; 24201 MW; F9B9F178A60CB3B4 CRC64;
 Query Match 1.5%; Score 8; DB 10; Length 224;
 Best Local Similarity 100.0%; Pred. No. 27;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 Qy 21 AAPAVEAE 28
 Db 84 AAPAVEAE 91
 RESULT 31
 O06077 PRELIMINARY; PRT; 229 AA.
 ID O06077;
 AC O06077;
 DT 01-JUL-1997 (TrEMBLrel. 04, Created)
 DT 01-JUL-1997 (TrEMBLrel. 04, Last sequence update)
 DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
 DE Hypothetical protein (Probable membrane protein).
 GN ML2615 OR MLCJ622.14.
 OS Mycobacterium leprae.
 OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
 OC Corynebacteriaceae; Mycobacteriaceae; Mycobacterium.
 OX NCBI_TaxID=1769;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=TN;
 RC MEDLINE=21128732; PubMed=11234002;
 RA Cole S.T., Biglmeier K., Parkhill J., James K.D., Thomson N.R.,
 RA Wheeler P.R., Honore N., Garnier T., Churcher C., Harris D.,
 RA Mungall K., Basham D., Brown D., Chillingworth T., Connor R.,
 RA Davies R.M., Devlin K., Duthoy S., Feltwell T., Fraser A., Hamlin N.,
 RA Holroyd S., Hornsby T., Jagels K., Lacroix C., Maclean J., Moule S.,
 RA Murphy L., Oliver K., Quail M.A., Rajandream M.A., Rutherford K.M.,
 RA Rutter S., Seeger K., Simon S., Simmonds M., Skelton J., Squares R.,
 RA Squares S., Stevens K., Taylor K., Whitehead S., Woodward J.R.,
 RA Barrell B.G.;
 RT "Massive gene decay in the leprosy bacillus.";
 RL Nature 409:1007-1011(2001).
 DR EMBL; Z95398; CAB08802.1; -;
 DR EMBL; AL583926; CAC32147.1; -;
 DR PIR; E87236; E87236.
 DR Leproma; ML2615; -;
 KW Hypothetical protein; Complete proteome.
 SQ SEQUENCE 229 AA; 24416 MW; 9A839BF68925E0A CRC64;
 Query Match 1.5%; Score 8; DB 16; Length 229;
 Best Local Similarity 100.0%; Pred. No. 28;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 Qy 18 PSQAAPAV 25
 Db 200 PSQAAPAV 207
 RESULT 32
 Q9R9Z0

Q8R9Z0 PRELIMINARY; PRT; 255 AA.
Q8R9Z0;
AC 01-JUN-2002 (TREMBlrel. 21, Created)
DT 01-JUN-2002 (TREMBlrel. 21, Last sequence update)
DT 01-OCT-2002 (TREMBlrel. 22, Last annotation update)
DE Flagellar biosynthesis/type III secretory pathway protein.
GN FLJH OR TT21440.
OS Thermoanaerobacter tengcongensis.
OC Bacteria; Firmicutes; Clostridia; Thermoanaerobacteriales;
OC Thermoanaerobacteriaceae; Thermoanaerobacter.
OX NCBI_TaxID=119072;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=MB4 / JCM 11007;
RX MEDLINE=21992816; PubMed=11997336;
RA Bao Q., Tian Y., Li W., Xu Z., Huang Z., Hu S., Dong W., Yang J.,
RA Chen Y., Xue Y., Xu Y., Lai X., Huang L., Dong X., Ma Y., Ling L.,
RA Tan H., Chen R., Wang J., Yu J., Yang H.;
RT "A complete sequence of T. tengcongensis genome.";
RL Genome Res. 12:689-700(2002).
DR EXBL; AB013102; BAM24662.1; -.
KW Complete Proteome.
SQ SEQUENCE 255 AA; 29724 MW; 9109A57784B8B859 CRC64;
Query Match 1.5%; Score 8; DB 16; Length 255;
Best Local Similarity 100.0%; Pred. No. 30;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 331 MLKAEER 338
Db |||||
75 MLKAEER 82
RESULT 33
Q8R9J9 PRELIMINARY; PRT; 255 AA.
AC 088P09;
DT 01-JUN-2003 (TREMBlrel. 24, Created)
DT 01-JUN-2003 (TREMBlrel. 24, Last sequence update)
DT 01-OCT-2003 (TREMBlrel. 25, Last annotation update)
DE Type IV pili biogenesis protein PilP.
GN PILP OR PP0851
OS Pseudomonas putida (strain KT2440).
OC Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales;
OC Pseudomonadaceae; Pseudomonas.
OX NCBI_TaxID=160488;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=22423060; PubMed=12534463;
RA Nelson K.B., Weinel C., Paulsen I.T., Dodson R.J., Hilbert H.,
RA Martins dos Santos V.A.P., Fouts D.E., Gill S.R., Pop M., Holmes M.,
RA Brinkac L., Beanan M., White O., Peterson J., Khouri H., Hance I.,
RA Madupu R., Nelson W., White O., Peterson J., Khouri H., Hance I.,
RA Chris Lee P., Holtzapfle E., Scanlan D., Tran K., Moazzez A.,
RA Uterback T., Rizzo M., Lee K., Kosack D., Moestl D., Wedler H.,
RA Lauber J., Stjepandic D., Hoheisel J., Straetz M., Heim S.,
RA Kiewitz C., Eisen J., Timmis K.N., Duesterhoeft A., Tuenmler B.,
RA Fraser C.M.;
RT "Complete genome sequence and comparative analysis of the
RT metabolically versatile Pseudomonas putida KT2440.";
RL Environ. Microbiol. 4:799-808(2002).
DR EMBL; AB016777; AAN66476.1; -.
DR TIGR; PP0851; -.
DR InterPro; IPR000437; Prok_lipoprot_S.
DR InterPro; IPR001440; TPR.
DR InterPro; IPR008941; TPR-like.
DR Pfam; PF00515; TPR; 4.
DR PROSITE; PS00013; PROKAR_LIPOPROTEIN; 1.
KW Complete proteome.
SQ SEQUENCE 255 AA; 28402 MW; C8FD220D181124FE CRC64;
Query Match 1.5%; Score 8; DB 16; Length 255;
Best Local Similarity 100.0%; Pred. No. 30;

Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 425 KALLEMAE 432
Db |||||
178 KALLEMAE 185
RESULT 34
Q86LQ1 PRELIMINARY; PRT; 264 AA.
AC 086LQ1;
DT 01-JUN-2003 (TREMBlrel. 24, Created)
DT 01-JUN-2003 (TREMBlrel. 24, Last sequence update)
DT 01-JUN-2003 (TREMBlrel. 24, Last annotation update)
DE Hypothetical muscle-derived protein.
OS Branchiostoma belcheri (Amphioxus).
OC Eukaryota; Metazoa; Chordata; Cephalochordata; Branchiostomidae;
OC Branchiostoma.
OX NCBI_TaxID=7741;
RN [1]
RP SEQUENCE FROM N.A.
RA Zhang H., Lin Y.;
RL Submitted (DEC-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; AV191780; AAO45170.1; -.
KW Hypothetical protein.
SQ SEQUENCE 264 AA; 31199 MW; EA0B82C96708B15B CRC64;
Query Match 1.5%; Score 8; DB 5; Length 264;
Best Local Similarity 100.0%; Pred. No. 31;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 373 LYTERFEE 380
Db |||||
92 LYTERFEE 99
RESULT 35
Q44174 PRELIMINARY; PRT; 281 AA.
AC 044174;
DT 01-JUN-1998 (TREMBlrel. 06, Created)
DT 01-JUN-1998 (TREMBlrel. 06, Last sequence update)
DT 01-OCT-2003 (TREMBlrel. 25, Last annotation update)
DE F58P6.1 protein.
GN F58P6.1
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC Rhabditidae; Peleoderinae; Caenorhabditis.
OX NCBI_TaxID=6239;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Bristol N2;
RX MEDLINE=99069613; PubMed=9851916;
RA None;
RT "Genome sequence of the nematode C. elegans: a platform for
RT investigating biology. The C. elegans Sequencing Consortium.";
RL Science 282:2012-2018(1998).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=Bristol N2;
RA Dante M., Kramer J., Gibson A.;
RT "The sequence of C. elegans cosmid F58P6.";
RL Submitted (DEC-1997) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=Bristol N2;
RA Waterston R.;
RL Submitted (MAY-1999) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF036699; AAB88358.1; -.
DR PIR; C88638; C88638.
DR WormPep; F58P6.1; CB17136.
DR GO; GO:0003824; F:catalytic activity; IEA.
DR GO; GO:0008152; P:metabolism; IEA.

DR InterPro; IPR008160; Collagen.
DR InterPro; IPR002486; Col_cuticle_N.
DR InterPro; IPR001150; Form_actrans_GR.
DR Pfam; PF01391; Collagen; 3.
DR Pfam; PF01484; Col_cuticle_N; 1.
DR PROSITE; PS00850; GLY_RADICAL; 1.
SQ SEQUENCE 281 AA; 27995 MW; EA2112F6CF753419 CRC64;

Query Match 1.5%; Score 8; DB 5; Length 281;
Best Local Similarity 100.0%; Pred. No. 33;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 77 GGPGEDEGA 84
Db 241 GGPGEDEGA 248
|||||
|||

RESULT 36
Q9SZW2 PRELIMINARY; PRT; 312 AA.
AC Q9SZW2;
DT 01-MAY-2000 (TREMBLrel. 13, Created)
DT 01-MAY-2000 (TREMBLrel. 12, Last sequence update)
DT 01-OCT-2003 (TREMBLrel. 25, Last annotation update)
DE Hypothetical protein.
GN F6G3.120 OR A74G30090.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids;
OC eurosids II; Brassicales; Brassicaceae; Arabidopsis.
OX NCBI_TaxID=3702;
RN [1]
RP SEQUENCE FROM N.A.
RA Bevan M., Murphy G., Ridley P., Hudson S., Bancroft I., Mewes H.W.,
RA Mayer K.F.X., Lemcke K., Schueller C.;
RL Submitted (MAY-1999) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RA EU Arabidopsis sequencing project;
RL Submitted (MAY-1999) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RA Murphy G., Ridley P., Hudson S., Mewes H.W., Lemcke K., Mayer K.F.X.;
RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
RN [4]
RP SEQUENCE FROM N.A.
RA EU Arabidopsis sequencing project;
RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AL078464; CAB43844.1; -;
DR EMBL; AL161576; CAB81002.1; -;
DR PIR; T08985; T08985
KW Hypothetical protein.
SQ SEQUENCE 312 AA; 35914 MW; D169444734E31331 CRC64;

Query Match 1.5%; Score 8; DB 10; Length 312;
Best Local Similarity 100.0%; Pred. No. 36;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 326 QOQOEWLX 333
Db 240 QOQOEWLX 247
|||||
|||

RESULT 37
Q8SA04 PRELIMINARY; PRT; 324 AA.
ID Q8SA04
AC Q8SA04;
DT 01-JUN-2002 (TREMBLrel. 21, Created)
DT 01-JUN-2002 (TREMBLrel. 21, Last sequence update)
DT 01-JUN-2003 (TREMBLrel. 24, Last annotation update)
DE Similar to root cap protein.
GN QJ1656 All.6.
OS Oryza Sativa (japonica cultivar-group).

OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
OC Ehrhartoideae; Oryzeae; Oryza.
OX NCBI_TaxID=39947;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=cv. Nipponbare;
RA Sasaki T., Matsumoto T., Yamamoto K.;
RT "Oryza sativa nipponbare (GA3) genomic DNA, chromosome 1, BAC
clone:QJ1656 All.6";
RL Submitted (MAR-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF003448; BAB85317.1; -;
DR Gramene; Q8SA04; -;
DR GO; GO:0016020; C:membrane; IEA.
DR GO; GO:0005215; P:transporter activity; IEA.
DR GO; GO:0006810; P:transport; IEA.
DR InterPro; IPR00515; BPD transp.
DR PROSITE; PS00402; BPD TRANSP INN MEMBER; 1.
SQ SEQUENCE 324 AA; 34945 MW; 0F37F890FD24EFF7 CRC64;

Query Match 1.5%; Score 8; DB 10; Length 324;
Best Local Similarity 100.0%; Pred. No. 38;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 494 AQPSSPR 501
Db 22 AQPSSPR 29
|||||
|||

RESULT 38
Q8ELJ3 PRELIMINARY; PRT; 346 AA.
ID Q8ELJ3
AC Q8ELJ3;
DT 01-MAR-2003 (TREMBLrel. 23, Created)
DT 01-MAR-2003 (TREMBLrel. 23, Last sequence update)
DT 01-OCT-2003 (TREMBLrel. 25, Last annotation update)
DE Hypothetical conserved protein.
GN OB3234.
OS Oceanobacillus iheyensis.
OC Bacteria; Firmicutes; Bacillales; Bacillaceae; Oceanobacillus.
OX NCBI_TaxID=182710;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=HT831 / DSM 14371 / JCM 11309;
RX MEDLINE=22220767; PubMed=12235376;
RA Takami H., Takaki Y., Uchiyama I.;
RT "Genome sequence of Oceanobacillus iheyensis isolated from the Iheya
RT Ridge and its unexpected adaptive capabilities to extreme
RT environments";
RL Nucleic Acids Res. 30:3927-3935(2002).
DR EMBL; AF004604; BAC15190.1; -;
DR GO; GO:0030288; C:periplasmic space (sensu Gram-negative Bact...; IEA.
DR InterPro; IPR000437; Prok_lipoprot_s.
DR InterPro; IPR005064; UPF0065.
DR Pfam; PF03401; UPF0065; 1.
DR PROSITE; PS00013; PROKAR_LIPOPROTEIN; 1.
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 346 AA; 37487 MW; 12B9A69D4807025A CRC64;

Query Match 1.5%; Score 8; DB 16; Length 346;
Best Local Similarity 100.0%; Pred. No. 40;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 379 EEPONTLS 386
Db 291 EEPONTLS 298
|||||
|||

RESULT 39
Q9Q1U7 PRELIMINARY; PRT; 348 AA.
ID Q9Q1U7
AC Q9Q1U7;
DT 01-MAY-2000 (TREMBLrel. 13, Created)

DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
 DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
 DE PF50b protein short form (Hypothetical protein).
 OS Saimiriine herpesvirus 2.
 OC Viruses; dsDNA viruses, no RNA stage; Herpesviridae;
 OC Gammaherpesvirinae; Rhadinovirus.
 OX NCBI_TaxID=10381;
 RN [1]
 RN SEQUENCE FROM N.A.
 RP STRAIN=C488;
 RX MEDLINE=20149233; PubMed=10683339;
 RA Thurau M., Whitehouse A., Wittmann S., Fickenscher H.;
 RA "Distinct transcriptional and functional properties of the R
 RT transactivator gene, orf50, of the transforming herpesvirus saimiri
 RT strain C488.";
 RL Virology 268:167-177 (2000).
 RN [2]
 RN SEQUENCE FROM N.A.
 RP STRAIN=C5753;
 RA Essner A., Thurau M., Wittmann S., Fickenscher H.;
 RA "Genomic sequence of herpesvirus saimiri C488 which is capable of
 RT transforming human T cells.";
 RT Submitted (APR-2001) to the EMBL/GenBank/DBJ databases.
 RN [3]
 RN SEQUENCE FROM N.A.
 RP STRAIN=C488;
 RC Fickenscher H., Thurau M., Wittmann S., Essner A.;
 RL Submitted (APR-2001) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AJ131935; CAC85421.1; -;
 DR EMBL; AJ410487; CAC84289.1; -;
 DR EMBL; AJ410493; CAC84346.1; -;
 DR GO; GO:0004386; F:Helicase activity; IEA.
 DR GO; GO:0009233; F:Peptidase activity; IEA.
 DR GO; GO:0016563; F:Transcriptional activator activity; IEA.
 DR GO; GO:0006355; P:Regulation of transcription, DNA-dependent; IEA.
 DR InterPro; IPR004998; Herpes_TAF50.
 DR PFam; PF03326; Herpes_TAF50; 1.
 KW Hypothetical protein; Alternating splicing; Helicase; Protease.
 SQ SEQUENCE 348 AA; 38472 MW; 8678581FAC735E8 CRC64;

Query Match 1.5%; Score 8; DB 12; Length 348;
 Best Local Similarity 100.0%; Pred. No. 40;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 151 GLGKEITL 158
 DB 2 GLGKEITL 9

RESULT 40
 Q805G5 PRELIMINARY; PRT; 348 AA.
 AC Q805G5;
 DT 01-JUN-2003 (TrEMBLrel. 24, Created)
 DT 01-JUN-2003 (TrEMBLrel. 24, Last sequence update)
 DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
 DE Hypothetical protein.
 OS Saimiriine herpesvirus 2.
 OC Viruses; dsDNA viruses, no RNA stage; Herpesviridae;
 OC Gammaherpesvirinae; Rhadinovirus.
 OX NCBI_TaxID=10381;
 RN [1]
 RN SEQUENCE FROM N.A.
 RP STRAIN=C5746, C5945, C5947, C5952, and C6661;
 RA Essner A., Thurau M., Wittmann S., Fickenscher H.;
 RA "Genomic sequence of herpesvirus saimiri C488 which is capable of
 RT transforming human T cells.";
 RT Submitted (APR-2001) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AJ410486; CAC85015.1; -;
 DR EMBL; AJ410488; CAC85021.1; -;
 DR EMBL; AJ410489; CAC85027.1; -;
 DR EMBL; AJ410490; CAC85033.1; -;
 DR EMBL; AJ410491; CAC85039.1; -;

DR GO; GO:0016563; P:transcriptional activator activity; IEA.
 DR GO; GO:0006355; P:regulation of transcription, DNA-dependent; IEA.
 DR InterPro; IPR004998; Herpes_TAF50.
 DR PFam; PF03326; Herpes_TAF50; 1.
 KW Hypothetical protein.
 SQ SEQUENCE 348 AA; 38449 MW; F512C3813CEID026 CRC64;
 Query Match 1.5%; Score 8; DB 12; Length 348;
 Best Local Similarity 100.0%; Pred. No. 40;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 151 GLGKEITL 158
 DB 2 GLGKEITL 9

RESULT 41
 Q9AES1 PRELIMINARY; PRT; 395 AA.
 ID Q9AES1
 AC Q9AES1;
 DT 01-JUN-2001 (TrEMBLrel. 17, Created)
 DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
 DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
 DE Major outer membrane protein OmpA.
 GN OMPA.
 OS Bacteroides fragilis.
 OC Bacteria; Bacteroidetes; Bacteroides (class); Bacteroidales;
 OC Bacteroidaceae; Bacteroides.
 OX NCBI_TaxID=817;
 RN [1]
 RN SEQUENCE FROM N.A.
 RC STRAIN=ATCC25285;
 RA Wexler H.M., Read E.K., Tomzynski T.J.;
 RT "Bacteroides fragilis (ATCC 25285) outer membrane protein OmpA (ompA)
 RT gene.";
 RL Submitted (MAR-2001) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AY028459; AAK20932.1; -;
 DR InterPro; IPR006664; Bac_OmpA.
 DR InterPro; IPR006665; OmpA/MotB.
 DR PFam; PF00691; OmpA; 1.
 DR PRINTS; PR01021; OMPADOMAIN.
 DR PRODOM; PR000930; OmpA/MotB; 1.
 SQ SEQUENCE 395 AA; 43347 MW; EFD85FC61034FBCA CRC64;

Query Match 1.5%; Score 8; DB 2; Length 395;
 Best Local Similarity 100.0%; Pred. No. 45;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 316 LQQQLVDA 323
 DB 263 LQQQLVDA 270

RESULT 42
 Q80BK0 PRELIMINARY; PRT; 405 AA.
 ID Q80BK0
 AC Q80BK0;
 DT 01-JUN-2003 (TrEMBLrel. 24, Created)
 DT 01-JUN-2003 (TrEMBLrel. 24, Last sequence update)
 DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
 DE Hypothetical protein.
 OS Saimiriine herpesvirus 2.
 OC Viruses; dsDNA viruses, no RNA stage; Herpesviridae;
 OC Gammaherpesvirinae; Rhadinovirus.
 OX NCBI_TaxID=10381;
 RN [1]
 RN SEQUENCE FROM N.A.
 RP STRAIN=A5747;
 RA Essner A., Thurau M., Wittmann S., Fickenscher H.;
 RA "Genomic sequence of herpesvirus saimiri C488 which is capable of
 RT transforming human T cells.";
 RT Submitted (APR-2001) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AJ410482; CAC84997.1; -;

```
DR GO: GO:0005634; C:nucleus; IEA.
DR GO: GO:0003677; F:DNA binding; IEA.
DR GO: GO:0016563; F:transcriptional activator activity; IEA.
DR GO: GO:0006355; P:regulation of transcription, DNA-dependent; IEA.
DR InterPro: IPR000637; AT_hoek.
DR InterPro: IPR004998; HePpes_TAF50.
DR Pfam: PF03326; Herpes_TAF50; 1.
DR PROSITE: PS00354; HMG1_Y; 1.
KW Hypothetical protein.
SQ SEQUENCE 405 AA; 44720 MW; 29CC90C991B61572 CRC64;

Query Match 1.5%; Score 8; DB 12; Length 405;
Best Local Similarity 100.0%; Pred. No. 46;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 151 GLGKEITL 158
Db 57 GLGKEITL 64

RESULT 43
ID Q805M7 PRELIMINARY; PRT; 405 AA.
AC Q805M7;
DT 01-JUN-2003 (TrEMBLrel. 24, Created)
DT 01-JUN-2003 (TrEMBLrel. 24, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Hypothetical protein.
OS Saimirine herpesvirus 2.
OC Viruses; dsDNA viruses, no RNA stage; Herpesviridae;
OC Gammaherpesvirinae; Rhadinovirus.
OX NCBI_TaxID=10381;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=A6355, and A6051;
RA Essner A., Thraun M., Wittmann S., Pickenscher H.;
RT "Genomic sequence of herpesvirus saimiri C488 which is capable of
transforming human T cells.";
RL Submitted (APR-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL: AJ410484; CAC84283.1; -.
DR EMBL: AJ410483; CAC85003.1; -.
DR GO: GO:0005634; C:nucleus; IEA.
DR GO: GO:0003677; F:DNA binding; IEA.
DR GO: GO:0016563; F:transcriptional activator activity; IEA.
DR GO: GO:0006355; P:regulation of transcription, DNA-dependent; IEA.
DR InterPro: IPR000637; AT_hoek.
DR InterPro: IPR004998; HePpes_TAF50.
DR Pfam: PF03326; Herpes_TAF50; 1.
DR PROSITE: PS00354; HMG1_Y; 1.
KW Hypothetical protein.
SQ SEQUENCE 405 AA; 44783 MW; 7512ED30F51045B7 CRC64;

Query Match 1.5%; Score 8; DB 12; Length 405;
Best Local Similarity 100.0%; Pred. No. 46;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 151 GLGKEITL 158
Db 57 GLGKEITL 64

RESULT 44
Q82DR9 PRELIMINARY; PRT; 413 AA.
ID Q82DR9;
AC Q82DR9;
DT 01-JUN-2003 (TrEMBLrel. 24, Created)
DT 01-JUN-2003 (TrEMBLrel. 24, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Hypothetical protein.
OS Streptomyces avermitilis.
OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC Streptomycineae; Streptomycetaceae; Streptomyces.

NCBI_TaxID=33903;
[1]
SEQUENCE FROM N.A.
STRAIN=MA-4680 / ATCC 31267 / NCIMB 12804 / NRRL 8165;
MEDLINE=21477403; PubMed=11572948;
RA Omura S., Ikeda H., Ishikawa J., Hanamoto A., Takahashi C.,
RA Shinose M., Takahashi Y., Horikawa H., Nakazawa H., Osonoe T.,
RA Kikuchi H., Shiba T., Sakaki Y., Hattori M.;
"Genome sequence of an industrial microorganism Streptomyces
RT avermitilis: deducing the ability of producing secondary
RT metabolites.";
RT Proc. Natl. Acad. Sci. U.S.A. 98:12215-12220(2001).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=MA-4680 / ATCC 31267 / NCIMB 12804 / NRRL 8165;
RX MEDLINE=22608306; PubMed=12692562;
RA Ikeda H., Ishikawa J., Hanamoto A., Shinose M., Kikuchi H., Shiba T.,
RA Sakaki Y., Hattori M., Omura S.;
"Complete genome sequence and comparative analysis of the industrial
RT microorganism Streptomyces avermitilis.";
RL Nat. Biotechnol. 21:526-531(2003).
DR EMBL: AP005040; BAC72611.1; -.
DR GO: GO:0016787; F:hydrolase activity; IEA.
DR InterPro: IPR006680; Amidohydro_1.
DR Pfam: PF01979; Amidohydro_1; 1.
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 413 AA; 41637 MW; 3B5CB2B80F14BDC0 CRC64;

Query Match 1.5%; Score 8; DB 16; Length 413;
Best Local Similarity 100.0%; Pred. No. 47;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 49 SGALRDVS 56
Db 297 SGALRDVS 304

RESULT 45
Q9L0M6 PRELIMINARY; PRT; 434 AA.
ID Q9L0M6;
AC Q9L0M6;
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE Putative hydrolase.
GN SCO4634 OR SCO82.05C.
OS Streptomyces coelicolor.
OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC Streptomycineae; Streptomycetaceae; Streptomyces.
OX NCBI_TaxID=1902;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=A3(2) / M145;
RX MEDLINE=21996410; PubMed=12000953;
RA Bentley S.D., Chater K.F., Cerdeno-Tarraga A.-M., Challis G.L.,
RA Thomson N.R., James K.D., Harris D.E., Quail M.A., Kieser H.,
RA Harper D., Bateman A., Brown S., Chandra G., Chen C.W., Collins M.,
RA Cronin A., Fraser A., Goble A., Hidalgo J., Hornsby T., Howarth S.,
RA Huang C.-H., Kieser T., Larke L., Murphy L., Oliver K., O'Neill S.,
RA Rabinowitsch E., Rajandream M.A., Rutherford K., Rutter S.,
RA Seeger K., Saunders D., Sharp S., Squares R., Squares S., Taylor K.,
RA Warren T., Wietzorrek A., Woodward J., Barrall B.G., Parkhill J.,
RA Hopwood D.A.;
"Complete genome sequence of the model actinomycete Streptomyces
RT coelicolor A3(2).";
RL Nature 417:141-147(2002).
DR EMBL: AL593120; CAB77408.1; -.
DR GO: GO:0016787; F:hydrolase activity; IEA.
DR InterPro: IPR006680; Amidohydro_1.
DR Pfam: PF01979; Amidohydro_1; 1.
KW Hydrolase; Complete proteome.
SQ SEQUENCE 434 AA; 43748 MW; P2BA6B51A465AB38 CRC64;
```

Query Match 1.5%; Score 8; DB 16; Length 434;
Best Local Similarity 100.0%; Pred. No. 49;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 49 SGALRDVS 56
ID 303 SGALRDVS 310
DB

RESULT 46
Q84VE3 PRELIMINARY; PRT; 436 AA.
AC Q84VE3;
DT 01-JUN-2003 (TREMELrel. 24, Created)
DT 01-JUN-2003 (TREMELrel. 24, Last sequence update)
DT 01-JUN-2003 (TREMELrel. 24, Last annotation update)
DE Muscle derived-like protein.
OS Eryza sativa (japonica cultivar-group).
OC Eukaryota; Viridiplantae; Streptophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
OC Ehrhartoideae; Oryzae; Oryza.
OX NCBI_TaxID=39347;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=cv. Nipponbare;
RA Cooper B., Hutchinson D., Park S., Guimil S., Luginbuhl P., Ellero C.,
RA Goff S., Glazebrook J.;
RT "Identification of Rice (Oryza sativa) Proteins Linked to the Cyclin-
RT Mediated Regulation of the Cell Cycle.";
RL Submitted (JAN-2003) to the EMBL/GenBank/DBJ databases.
DR EMBL; AY224541; AA072661.1; -- 9180958F88E2328 CRC64;
SQ SEQUENCE 436 AA; 48148 MW; 9180958F88E2328 CRC64;

Query Match 1.5%; Score 8; DB 10; Length 436;
Best Local Similarity 100.0%; Pred. No. 49;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 224 KLESICRE 231
DB 186 KLESICRE 193

RESULT 47
Q80BH6 PRELIMINARY; PRT; 439 AA.
AC Q80BH6;
DT 01-JUN-2003 (TREMELrel. 24, Created)
DT 01-JUN-2003 (TREMELrel. 24, Last sequence update)
DT 01-OCT-2003 (TREMELrel. 25, Last annotation update)
DE Hypothetical protein.
OS Saimirine herpesvirus 2.
OC Viruses; GSDNA viruses, no RNA stage; Herpesviridae;
OC Gammaherpesvirinae; Rhadinovirus.
OX NCBI_TaxID=10381;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CS907;
RA Ensser A., Thurau M., Wittmann S., Fickenscher H.;
RT "Genomic sequence of herpesvirus saimiri C488 which is capable of
RT transforming human T cells.";
RL Submitted (APR-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AJ410492; CAC85045.1; --
DR GO; GO:0016563; F:transcriptional activator activity; IEA.
DR GO; GO:0006355; P:regulation of transcription, DNA-dependent; IEA.
DR InterPro; IPR004998; Herpes_TAF50.
DR Pfam; PF03326; Herpes_TAF50; 1.
KW Hypothetical protein.
SQ SEQUENCE 439 AA; 49007 MW; 2643B9384E505B2D CRC64;

Query Match 1.5%; Score 8; DB 12; Length 439;
Best Local Similarity 100.0%; Pred. No. 50;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 151 GLGKEITL 158
DB 93 GLGKEITL 100

RESULT 48
Q9ZLG8 PRELIMINARY; PRT; 450 AA.
AC Q9ZLG8;
DT 01-MAY-1999 (TREMELrel. 10, Created)
DT 01-MAY-1999 (TREMELrel. 10, Last sequence update)
DT 01-OCT-2003 (TREMELrel. 25, Last annotation update)
DE Putative.
GN JHP0612.
OS Helicobacter pylori J99 (Campylobacter pylori J99).
OC Bacteria; Proteobacteria; Epsilonproteobacteria; Campylobacterales;
OC Helicobacteraceae; Helicobacter.
OX NCBI_TaxID=85963;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=99120557; PubMed=9923682;
RA Alm R.A., Ling L.-S.L., Moir D.T., Kirg B.L., Brown E.D., Doig P.C.,
RA Smith D.R., Noonan B., Guild B.C., deJonge B.L., Carmel G.,
RA Tummino P.J., Caruso A., Uria-Nickelsen M., Mills D.M., Ives C.,
RA Gibson R., Merberg D., Mills S.D., Jiang Q., Taylor D.E., Vovis G.F.,
RA Trust T.J.;
RT "Genomic sequence comparison of two unrelated isolates of the human
RT Gastric pathogen Helicobacter pylori.";
RL Nature 387:176-180(1999).
DR EMBL; AB001493; AAD06193.1; --
DR PIR; E71909; E71909.
DR GO; GO:0005524; F:ATP binding; IEA.
DR GO; GO:0008026; F:ATP dependent helicase activity; IEA.
DR InterPro; IPR001410; DEAD.
DR InterPro; IPR007560; Mr_Cat.
DR Pfam; PF04471; Mr_Cat; 1.
DR SMART; SM00487; DEXDc; 1.
KW Complete proteome.
SQ SEQUENCE 450 AA; 51347 MW; E15CFADF2B9074C8 CRC64;

Query Match 1.5%; Score 8; DB 16; Length 450;
Best Local Similarity 100.0%; Pred. No. 51;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 63 LEDILSTY 70
DB 283 LEDILSTY 290

RESULT 49
Q8JZS6 PRELIMINARY; PRT; 451 AA.
AC Q8JZS6;
DT 01-MAR-2003 (TREMELrel. 23, Created)
DT 01-MAR-2003 (TREMELrel. 23, Last sequence update)
DT 01-MAR-2003 (TREMELrel. 23, Last annotation update)
DE Hypothetical protein (Fragment).
OS Gibberella fujikuroi (Bakanae and foot rot disease fungus) (Fusarium moniliforme).
OC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;
OC Hypocreomycetidae; Hypocreales; Nectriaceae; Gibberella.
OX NCBI_TaxID=5127;
RN [1]
RP SEQUENCE FROM N.A.
RA Yoshida H., Iizuka M., Hirabayashi T.;
RT "Cloning and sequencing of a cDNA encoding caldesmon-like protein of
RT Gibberella fujikuroi.";
RL Submitted (SEP-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; AB023443; BAC16636.1; --
KW Hypothetical protein.
FT NON TER
SQ SEQUENCE 451 AA; 53407 MW; 0D2796E91E55D610 CRC64;

```

Query Match      1.5%; Score 8; DB 3; Length 451;
Best Local Similarity 100.0%; Pred. No. 51;
Matches      8; Conservative      0; Mismatches      0; Indels      0; Gaps      0;

QY      248 REEEERK 255
Db      15 REEEERK 22

RESULT 50
ID Q8NMB0 PRELIMINARY; PRT; 484 AA.
AC Q8NMB0;
DT 01-OCT-2002 (TrEMBLrel. 22, Created)
DT 01-OCT-2002 (TrEMBLrel. 22, Last sequence update)
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE NAD-dependent aldehyde dehydrogenases (EC 1.2.1.-).
CGL2668.
GN Corynebacterium glutamicum (Brevibacterium flavum).
OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC Corynebacterineae; Corynebacteriaceae; Corynebacterium.
OX NCBI_TaxID=1718;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 13032 / DSM 20300 / NCIB 10025;
RA Nakagawa S.;
RL "Complete genomic sequence of Corynebacterium glutamicum ATCC 13032.";
DR EMBL: AP005282; BAC00062.1; -.
DR GO; GO:0016491; P:oxidoreductase activity; IEA.
DR GO; GO:0008152; P:metabolism; IEA.
DR InterPro: IPR02086; Aldehyde_dehydr.
DR Pfam: PF00171; aldedh; 1.
DR PROSITE: PS00687; ALDEHYDE DEHYDR GLU; 1.
DR Oxidoreductase; Complete proteome.
KM Oxidoreductase; Complete proteome.
SQ SEQUENCE 484 AA; 51084 MW; A2E416EB585775F CRC64;

Query Match      1.5%; Score 8; DB 16; Length 484;
Best Local Similarity 100.0%; Pred. No. 54;
Matches      8; Conservative      0; Mismatches      0; Indels      0; Gaps      0;

QY      181 AELEEHR 188
Db      80 AELEEHR 87

RESULT 51
ID Q8PEY9 PRELIMINARY; PRT; 501 AA.
AC Q8PEY9;
DT 01-OCT-2002 (TrEMBLrel. 22, Created)
DT 01-OCT-2002 (TrEMBLrel. 22, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Polyvinylalcohol dehydrogenase.
GN XAC4199.
OC Xanthomonas axonopodis (pv. citri).
OC Bacteria; Proteobacteria; Gammaproteobacteria; Xanthomonadales;
OC Xanthomonadaceae; Xanthomonas.
OX NCBI_TaxID=92829;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=306 / ATCC 13902 / XV 101;
RA MEDLINE=2202145; PubMed=12024217;
RA da Silva A.C.R., Ferro J.A., Reinach F.C., Farah C.S., Furlan L.R.,
RA Alves L.M.C., do Amaral A.M., Bertolini M.C., Camargo L.E.A.,
RA Quaggio R.B., Monteiro-Vicorello C.B., van Sluys M.A., Almeida N.P.,
RA Canarotte G., Cannavini F., Cardoso J., Chamargo F., Ciapina L.P.,
RA Cicarelli R.M.B., Coutinho L.L., Cursino-Santos J.R., El-Dorry H.,
RA Faria J.B., Ferreira A.J.S., Ferreira R.C.C., Ferro M.I.T.,
RA Formighieri E.F., Franco M.C., Greggio C.C., Gruber A.,
RA Katsuyama A.M., Kishi E.T., Leite R.P., Lemos E.G.M., Lemos M.V.F.,
RA Locall E.C., Machado M.A., Madeira A.M.B.N., Martinez-Rossi N.M.,
RA Martins E.C., Meidanis J., Menck C.F.N., Miyaki C.Y., Moon D.H.,

```

```

RA Moreira L.M., Novo M.T.M., Okura V.X., Oliveira M.C., Oliveira V.R.,
RA Pereira H.A., Rossi A., Sena J.A.D., Silva C., de Souza R.F.,
RA Spínola L.A.F., Takita M.A., Tamura R.E., Teixeira E.C., Tezsa R.I.D.,
RA Trindade dos Santos M., Truffi D., Tsai S.W., White F.F.,
RA Setubal J.C., Kitajima J.P.;
RT "Comparison of the genomes of two Xanthomonas pathogens with differing
RT host specificities.";
RL Nature 417:459-463(2002).
DR EMBL: AB012071; BAM39034.1; -.
DR GO; GO:0006118; P:electron transport; IEA.
DR InterPro: IPR002372; Bac_PQQ_repeat.
DR Pfam: PF01011; PQQ; 2.
KW Complete proteome.
SQ SEQUENCE 501 AA; 52817 MW; F21C9784E3652129 CRC64;

Query Match      1.5%; Score 8; DB 16; Length 501;
Best Local Similarity 100.0%; Pred. No. 56;
Matches      8; Conservative      0; Mismatches      0; Indels      0; Gaps      0;

QY      474 AGGQGSLT 481
Db      417 AGGQGSLT 424

RESULT 52
ID Q9QIU8 PRELIMINARY; PRT; 520 AA.
AC Q9QIU8;
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE ORF50a protein long form (Transcriptional regulator).
OS Salmirine herpesvirus 2.
OC Viruses; dsDNA viruses, no RNA stage; Herpesviridae;
OC Gammaherpesvirinae; Rhadinovirus.
OX NCBI_TaxID=10381;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C488;
RX MEDLINE=20149233; PubMed=10683339;
RA Thurauf M., Whitehouse A., Wittmann S., Meredith D., Fickenscher H.;
RT "Distinct transcriptional and functional properties of the R
RT transactivator gene, orf50, of the transforming herpesvirus saimiri
RT strain C488.";
RL Virology 268:167-177(2000).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=C488;
RA Fickenscher H., Thurauf M., Wittmann S., Ensser A.;
RL Submitted (APR-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL: AJ131935; CAB65420.1; -.
DR BMSL: AJ410493; CAC84344.1; -.
DR GO; GO:0004386; P:helicase activity; IEA.
DR GO; GO:0008233; P:peptidase activity; IEA.
DR GO; GO:0016563; P:transcriptional activator activity; IEA.
DR GO; GO:0006355; P:regulation of transcription, DNA-dependent; IEA.
DR InterPro: IPR004998; Herpes_TAF50.
DR Pfam: PF03326; Herpes_TAF50.1.
KW Alternative splicing; Helicase; Protease.
SQ SEQUENCE 520 AA; 58309 MW; 33CADF47EF50AD5D CRC64;

Query Match      1.5%; Score 8; DB 12; Length 520;
Best Local Similarity 100.0%; Pred. No. 58;
Matches      8; Conservative      0; Mismatches      0; Indels      0; Gaps      0;

QY      151 GLGKEITL 158
Db      174 GLGKEITL 181

RESULT 53
Q8VCQ8 PRELIMINARY; PRT; 530 AA.
ID Q8VCQ8

```

AC Q8VC08;
DT 01-MAR-2002 (TrEMBLrel. 20, Created)
DT 01-MAR-2002 (TrEMBLrel. 20, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Similar to caldesmon 1.
GN CALDI OR 4833423D12RIK.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Cranialata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Breast tumor;
RA Strausberg K.;
RL Submitted (DEC-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; BC019435; ARI19435.1; -.
DR MGD; MGI:88250; Caldi.
DR GO; GO:0003779; F:actin binding; IEA.
DR GO; GO:0005516; F:calmodulin binding; IEA.
DR GO; GO:0017022; F:myosin binding; IEA.
DR GO; GO:0006936; F:muscle contraction; IEA.
DR InterPro; IPR006017; Caldesmon.
DR InterPro; IPR006018; Caldesmon_LSP.
DR Pfam; PF02029; Caldesmon; 1.
DR PRINTS; PR01076; CALDESMON.
SQ SEQUENCE 530 AA; 50453 MW; 697BFACF8FD68A2E CRC64;
Query Match 1.5%; Score 8; DB 11; Length 530;
Best Local Similarity 100.0%; Pred. No. 59;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 247 AREEEKR 254
DB 333 AREEEKR 340
RESULT 54
Q8P3J1 PRELIMINARY; PRT; 539 AA.
ID Q8P3J1
AC Q8P3J1
DT 01-OCT-2002 (TrEMBLrel. 22, Created)
DT 01-OCT-2002 (TrEMBLrel. 22, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Polyvinylalcohol dehydrogenase.
GN XCC4080.
OS Xanthomonas campestris (pv. campestris).
OC Bacteria; Proteobacteria; Gammaproteobacteria; Xanthomonadales;
OC Xanthomonadaceae; Xanthomonas.
OX NCBI_TaxID=340;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 33913 / NCPPB 528;
RX MEDLINE=22022145; PubMed=12024217;
RA da Silva A.C.R., Ferro J.A., Reinach F.C., Parah C.S., Furlan L.R.,
RA Quaggio R.B., Monteiro-Vitorello C.B., Van Sluys M.A., Almeida N.F.,
RA Alves L.M.C., do Amaral A.M., Bertolini M.C., Camargo L.E.A.,
RA Camarotte G., Canavan F., Cardozo J., Chamberg F., Chapina L.P.,
RA Cicarelli R.M.B., Coutinho L.L., Cursino-Santos J.R., El-Dorri H.,
RA Faria J.B., Ferreira A.J.S., Ferreira R.C.C., Ferro M.I.T.,
RA Formighieri E.F., Franco M.C., Greggio C.C., Gruber A.,
RA Katsuyama A.M., Kishi L.T., Leite R.P., Lemos E.G.M., Lemos M.V.P.,
RA Locali E.C., Machado M.A., Madeira A.M.B.N., Martinez-Rossi N.M.,
RA Martins E.C., Meidanis J., Menck C.F.M., Miyaki C.Y., Moon D.H.,
RA Moreira L.M., Novo M.T.M., Okura V.K., Oliveira M.C., Oliveira V.R.,
RA Pereira L.A., Rossi A., Sena J.A.D., Silva C., de Souza R.F.,
RA Spínola L.A.F., Takita M.A., Tamura E.C., Teixeira R.I.D.,
RA Trindade dos Santos M., Truffi D., Tsai S.M., White F.P.,
RA Setubal J.C., Kitajima J.P.;
RT "Comparison of the genomes of two Xanthomonas pathogens with differing
RT host specificities."
RL Nature 417:459-463 (2002).
DR EMBL; AR012531; XAM43301.1; -.
DR GO; GO:0005524; F:ATP binding; IEA.

DR GO; GO:0004812; F:RNA ligase activity; IEA.
DR GO; GO:0006418; P:amino acid activation; IEA.
DR GO; GO:0006118; P:electron transport; IEA.
DR InterPro; IPR002372; Bac_PQQ_repeat.
DR InterPro; IPR001412; tRNA-synt_1.
DR Pfam; PF01011; PQQ; 2
DR PROSITE; PS00178; AA-TFNA_LIGASE_I; 1.
KW Complete proteome.
SQ SEQUENCE 539 AA; 56714 MW; A47CA59B5F43BF12 CRC64;
Query Match 1.5%; Score 8; DB 16; Length 539;
Best Local Similarity 100.0%; Pred. No. 60;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 474 AGGQSLT 481
DB 455 AGGQSLT 462
RESULT 55
Q93ZM4 PRELIMINARY; PRT; 558 AA.
ID Q93ZM4
AC Q93ZM4;
DT 01-DEC-2001 (TrEMBLrel. 19, Created)
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE AT3G48870/T21J18.140.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids;
OC eurosids II; Brassicales; Brassicaceae; Arabidopsids.
OX NCBI_TaxID=3702;
RN [1]
RP SEQUENCE FROM N.A.
RA Cheuk R., Chen H., Kim C.J., Koesema E., Meyers M.C., Banh J.,
RA Bowser L., Carninci P., Dale J.M., Goldsmith A.D., Hayashizaki Y.,
RA Ishida J., Jiang P.X., Jones T., Kamiya A., Karlin-Neumann G.,
RA Kawai J., Lam B., Lee J.M., Lin J., Liu S.X., Miranda M., Narusaka M.,
RA Nguyen M., Onodera C.S., Palm C.J., Pham P.K., Quach H.L., Sakurai T.,
RA Satou M., Seki M., Southwick A., Tang C.C., Toriumi M., Yamada K.,
RA Yamamura Y., Yu G., Yu S., Shinozaki K., Davis R.W., Theologis A.,
RA Ecker J.R.;
RT "Arabidopsis cDNA clones";
RL Submitted (SEP-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AY056787; AAL10478.1; -.
DR GO; GO:0005524; F:ATP binding; IEA.
DR GO; GO:0003754; F:chaperone activity; IEA.
DR GO; GO:0000166; F:nucleotide binding; IEA.
DR InterPro; IPR003593; AAA_ATPase.
DR InterPro; IPR003959; AAA_ATPase_centr.
DR InterPro; IPR001270; Chaprin_clipA/B.
DR Pfam; PF00004; AAA; 1.
DR Pfam; PF02861; Clp_N; 3.
DR SMART; SM00382; AAA; 1.
DR PROSITE; PS00870; CLPAB_1; 1.
KW ATP-binding.
SQ SEQUENCE 558 AA; 61686 MW; 7B96AB5B1B57221 CRC64;
Query Match 1.5%; Score 8; DB 10; Length 558;
Best Local Similarity 100.0%; Pred. No. 62;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 127 BEIRQSD 134
DB 381 BEIRQSD 388
RESULT 56
Q97WX9 PRELIMINARY; PRT; 568 AA.
ID Q97WX9
AC Q97WX9;
DT 01-OCT-2001 (TrEMBLrel. 18, Created)

01-OCT-2001 (TREMELrel. 19, Last sequence update)
01-OCT-2003 (TREMELrel. 25, Last annotation update)
DE Hypothetical protein, plasmid pNOB8 orfs 620 and 630A
DE homolog.
GN SSO1981.
OS Sulfolobus solfataricus.
OC Archaea; Crenarchaeota; Thermoprotei; Sulfolobales; Sulfolobaceae;
OC Sulfolobus.
OX NCBI_TaxID=2287;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 35092 / DSM 1617 / P2;
RX MEDLINE=21332726; PubMed=11427726;
RA She Q., Singh R.K., Confalonieri F., Zivanovic Y., Allard G.,
RA Aways M.J., Chan-Weiler C.-Y., Clausen I.G., Curtis B.A.,
RA De Moors A., Erauso G., Fletcher C., Gordon P.M.K.,
RA Heikamp-de Jong I., Jeffries A.C., Kozera C.J., Medina N., Peng X.,
RA Thi-Ngoc H.P., Redder P., Schenk M.B., Theriault C., Tolstrup N.,
RA Charlebois R.L., Doolittle W.P., Duguet M., Gaasterland T.,
RA Garrett R.A., Ragan M.A., Senses C.W., Van der Oost J.,
RT "the complete genome of the crenarchaeon Sulfolobus solfataricus P2.";
RL Proc. Natl. Acad. Sci. U.S.A. 98:7835-7840(2001).
DR EMBL; AF006805; AAK42172.1; -;
DR PIR; E90364; E90364.
DR InterPro; IPR003016; Lipoyl BS.
DR PROSITE; PS00189; LIPOYL; 1.
KK Hypothetical protein; Complete proteome.
SQ SEQUENCE 568 AA; 63993 MW; D32B40D02D097985 CRC64;

Query Match 1.5%; Score 8; DB 17; Length 568;
Best Local Similarity 100.0%; Pred. No. 63;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 404 KKIKLEK 411
DB 91 KKIKLEK 98

RESULT 57
Q8DJCO PRELIMINARY; PRT; 609 AA.
ID Q8DJCO
AC Q8DJCO
DT 01-MAR-2003 (TREMELrel. 23, Created)
DT 01-MAR-2003 (TREMELrel. 23, Last sequence update)
DT 01-JUN-2003 (TREMELrel. 24, Last annotation update)
DE Precorrin methylase.
GN COBJ OR TLL1307.
OS Synechococcus elongatus (Thermosynechococcus elongatus).
OC Bacteria; Cyanobacteria; Chroococcales; Synechococcus.
OX NCBI_TaxID=32046;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=BP-1;
RX MEDLINE=2225144; PubMed=12240834;
RA Nakamura Y., Kaneko T., Sato S., Ikeuchi M., Katoh H., Sasamoto S.,
RA Watanabe A., Iriguchi M., Kawashima K., Kimura T., Kishida Y.,
RA Kiyokawa C., Kohara M., Matsumoto M., Matsumoto A., Nakazaki N.,
RA Shimp S., Sugimoto M., Takeuchi C., Yamada M., Tabata S.,
RT "Complete genome structure of the thermophilic cyanobacterium
Thermosynechococcus elongatus BP-1.";
RL DNA Res. 9:123-130(2002).
DR EMBL; AP005373; BAC08859.1; -;
DR GO; GO:0008168; F:methyltransferase activity; IEA.
DR GO; GO:0008152; P:metabolism; IEA.
DR GO; GO:0009236; P:vitamin B12 biosynthesis; IEA.
DR InterPro; IPR002750; CbiG.
DR InterPro; IPR006363; Cobj.
DR InterPro; IPR000878; Cor/por_Mettransf.
DR Pfam; PF01890; CbiG; 1.
DR Pfam; PF00590; TP_methylase; 1.
DR TIGRPFams; TIGR01466; cobJ_cbiH; 1.
KW Methyltransferase; Complete proteome.
SQ SEQUENCE 609 AA; 54732 MW; 2A26F93FB24B920D CRC64;

Query Match 1.5%; Score 8; DB 16; Length 609;
Best Local Similarity 100.0%; Pred. No. 67;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 317 QQQLVDK 324
DB 532 QQQLVDK 539

RESULT 58
Q98P27 PRELIMINARY; PRT; 649 AA.
ID Q98P27
AC Q98P27
DT 01-OCT-2001 (TREMELrel. 18, Created)
DT 01-OCT-2001 (TREMELrel. 18, Last sequence update)
DT 01-JUN-2003 (TREMELrel. 24, Last annotation update)
DE Hypothetical protein ml19638.
GN ML19638.
OS Rhizobium loti (Mesorhizobium loti).
OG Plasmid pMLB.
OC Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;
OC Phyllobacteriaceae; Mesorhizobium.
OX NCBI_TaxID=381;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=MAFF303099;
RX MEDLINE=21082930; PubMed=11214968;
RA Kaneko T., Nakamura Y., Sato S., Asamizu E., Kato T., Sasamoto S.,
RA Watanabe A., Ideawara K., Ishikawa A., Kawashima K., Kimura T.,
RA Kishida Y., Kiyokawa C., Kohara M., Matsumoto M., Matsumoto A.,
RA Mochizuki Y., Nakayama S., Nakazaki N., Shimp S., Sugimoto M.,
RA Takeuchi C., Yamada M., Tabata S.,
RT "Complete genome structure of the nitrogen-fixing symbiotic bacterium
Rhizobium loti.";
RL DNA Res. 7:331-338(2000).
DR EMBL; AP003017; BAB54828.1; -;
DR GO; GO:0046821; C:extrachromosomal DNA; IEA.
KW Plasmid; Hypothetical protein; Complete proteome.
SQ SEQUENCE 649 AA; 74481 MW; D8D49A4A5B2998 CRC64;

Query Match 1.5%; Score 8; DB 16; Length 649;
Best Local Similarity 100.0%; Pred. No. 71;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 289 LAERLKL 296
DB 124 LAERLKL 131

RESULT 59
Q92QD4 PRELIMINARY; PRT; 649 AA.
ID Q92QD4
AC Q92QD4
DT 01-DEC-2001 (TREMELrel. 19, Created)
DT 01-DEC-2001 (TREMELrel. 19, Last sequence update)
DT 01-MAR-2002 (TREMELrel. 20, Last annotation update)
DE Hypothetical protein R01397.
GN R01397 OR SMC01267.
OS Rhizobium meliloti (Sinorhizobium meliloti).
OC Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;
OC Rhizobiaceae; Sinorhizobium/Ensifer group; Sinorhizobium.
OX NCBI_TaxID=382;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=1021;
RX MEDLINE=21596507; PubMed=11481430;
RA Capela D., Barloy-Hubier F., Gouzy J., Bothe G., Ampe F., Batut J.,
RA Boistard P., Becker A., Boutery M., Cadieu E., Dreano S., Gloux S.,
RA Godrie T., Goffeau A., Kahn D., Kiss E., Lelaure V., Masuy D.,
RA Pohl T., Portetelle D., Puehler A., Purnelle B., Ramsberger U.,
RA Renard C., Thebault P., Vandenbol M., Weidner S., Galibert F.,
RT "Analysis of the chromosome sequence of the legume symbiont

```

RT Sinorhizobium meliloti strain 1021.;
RL Proc. Natl. Acad. Sci. U.S.A. 98:9877-9882(2001).
DR EMBL: AL591787; CAC45976.1; -.
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 649 AA; 74239 MW; 8578F0EA569721D5 CRC64;

Query Match 1.5%; Score 8; DB 16; Length 649;
Best Local Similarity 100.0%; Pred. No. 71;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 289 LAERLUKL 296
DB 124 LAERLUKL 131

RESULT 60
O48931 PRELIMINARY; PRT; 928 AA.
AC O48931
DT 01-JUN-1998 (TrEMBLrel. 06, Created)
DT 01-JUN-1998 (TrEMBLrel. 06, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE ClpC.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids;
OC eurosid II; Brassicales; Brassicaceae; Arabidopsi.
OX NCBI_TaxID=3702;
RN [1]
RP SEQUENCE FROM N.A.
RA Shinn P., Chen H., Cheuk R., Kim C.J., Meyers M.C., Sanh J.,
RA Bowser L., Carninci P., Chang E., Dale J.M., Goldsmith A.D.,
RA Hayashizaki Y., Ishida J., Jones T., Kamiya A., Karlin-Neumann G.,
RA Kawai J., Lam B., Lee J.M., Lin J., Miranda M., Narusaka M.,
RA Nguyen M., Onodera C.S., Palm C.J., Quach H.L., Sakurai T., Satou M.,
RA Seki M., Southwick A., Tang C.C., Toriumi M., Wu H.C., Yamada K.,
RA Yamamura Y., Yu G., Yu S., Shinozaki K., Davis R.W., Theologis A.,
RA Ecker J.R.;
RT "Arabidopsis cDNA clones.";
RL Submitted (MAY-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL: AB017063; BAB08738.1; -.
DR EMBL: AY102125; AAM26692.1; -.
DR GO: GO:0005524; F:ATP binding; IEA.
DR GO: GO:0003754; F:chaperone activity; IEA.
DR GO: GO:0003677; F:DNA binding; IEA.
DR GO: GO:0004518; F:nuclease activity; IEA.
DR GO: GO:0006289; F:nucleotide-excision repair; IEA.
DR InterPro: IPR003959; AAA ATPase centr.
DR InterPro: IPR001270; Chaprinin_clpA/B.
DR InterPro: IPR004176; Clp N.
DR InterPro: IPR001943; UvrE/C.
DR Pfam: PF00004; AAA; 2.
DR Pfam: PF02861; Clp_N; 2.
DR PRINTS: PR00300; CLPPTPTEASEA.
DR PROSITE: PS00870; CLPAB_1; 1.
DR PROSITE: PS00871; CLPAB_2; 1.
DR PROSITE: PSS0151; UVR; 1.
KW ATP-binding; Protease.
SQ SEQUENCE 928 AA; 103455 MW; 103452 MW; 438DEA514125F0BF CRC64;

Query Match 1.5%; Score 8; DB 10; Length 928;
Best Local Similarity 100.0%; Pred. No. 98;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 127 EIRQSD 134
DB 360 EIRQSD 367

RESULT 62
Q9UP82 PRELIMINARY; PRT; 944 AA.
AC Q9UP82
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT 01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
DE Mucin 4 (Fragment).
GN MUC4
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;

ATP-dependent Clp protease, ATP-binding subunit
(AT5950920/K3X7.7).
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids;
OC eurosid II; Brassicales; Brassicaceae; Arabidopsi.
OX NCBI_TaxID=3702;
RN [1]
RP SEQUENCE FROM N.A.
RA Kaneko T., Katoh T., Sato S., Nakamura Y., Asamizu E., Kotani H.,
RA Miyajima N., Tabata S.;
RT "Structural analysis of Arabidopsis thaliana chromosome 5. IX.
RT Sequence features of the regions of 1,011,550 bp covered by seventeen
RT P1 and TAC clones.";
RL DNA Res. 6:183-195(1999).
RN [2]
RP SEQUENCE FROM N.A.
RA Shinn P., Chen H., Cheuk R., Kim C.J., Meyers M.C., Sanh J.,
RA Bowser L., Carninci P., Chang E., Dale J.M., Goldsmith A.D.,
RA Hayashizaki Y., Ishida J., Jones T., Kamiya A., Karlin-Neumann G.,
RA Kawai J., Lam B., Lee J.M., Lin J., Miranda M., Narusaka M.,
RA Nguyen M., Onodera C.S., Palm C.J., Quach H.L., Sakurai T., Satou M.,
RA Seki M., Southwick A., Tang C.C., Toriumi M., Wu H.C., Yamada K.,
RA Yamamura Y., Yu G., Yu S., Shinozaki K., Davis R.W., Theologis A.,
RA Ecker J.R.;
RT "Arabidopsis cDNA clones.";
RL Submitted (MAY-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL: AB017063; BAB08738.1; -.
DR EMBL: AY102125; AAM26692.1; -.
DR GO: GO:0005524; F:ATP binding; IEA.
DR GO: GO:0003754; F:chaperone activity; IEA.
DR GO: GO:0003677; F:DNA binding; IEA.
DR GO: GO:0004518; F:nuclease activity; IEA.
DR GO: GO:0006289; F:nucleotide-excision repair; IEA.
DR InterPro: IPR003959; AAA ATPase centr.
DR InterPro: IPR001270; Chaprinin_clpA/B.
DR InterPro: IPR004176; Clp N.
DR InterPro: IPR001943; UvrE/C.
DR Pfam: PF00004; AAA; 2.
DR Pfam: PF02861; Clp_N; 2.
DR PRINTS: PR00300; CLPPTPTEASEA.
DR PROSITE: PS00870; CLPAB_1; 1.
DR PROSITE: PS00871; CLPAB_2; 1.
DR PROSITE: PSS0151; UVR; 1.
KW ATP-binding; Protease.
SQ SEQUENCE 928 AA; 103455 MW; 103452 MW; 438DEA514125F0BF CRC64;

Query Match 1.5%; Score 8; DB 10; Length 928;
Best Local Similarity 100.0%; Pred. No. 98;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 127 EIRQSD 134
DB 359 EIRQSD 366

RESULT 61
Q9FI56 PRELIMINARY; PRT; 929 AA.
AC Q9FI56
DT 01-MAR-2001 (TrEMBLrel. 16, Created)
DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)

```

(1)
RP MEDLINE=99077625; PubMed=9858486;
Gipson I.K., Spurr-Michaud S., Moccia R., Zhan Q., Toribara N.,
Ho S.B., Gargiulo A.R., Hill J.A. III;
"MUC4 and MUC5B transcripts are the prevalent
ribonucleic acids of the human endocervix";
Biol. Reprod. 60:58-64(1999).
EMBL: AF058803; AAC34750.1; -.
NON_TER 1 1
NON_FT 944 944
SEQUENCE 944 AA; 96314 MW; 5B87CE9EC5085535 CRC64;

Query Match 1.5%; Score 8; DB 4; Length 944;
Best Local Similarity 100.0%; Pred.No. 99;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 513 STEASQT 520
Db 697 STEASQT 704

RESULT 63
Q9MZ26 PRELIMINARY; PRT; 952 AA.
AC Q9MZ26;
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE AtClpC.
OS T21J18.140.
GN Arabidopsis thaliana (Mouse-ear cross).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids;
OC eurosoids II; Brassicales; Brassicaceae; Arabidopsi.
OX NCBI_TaxID=3702;
RN [1]
RP SEQUENCE FROM N.A.
RA Rieger M., Gabel C., Mueller-Auer S., Schaefer M., Zipp M.,
RA Mewes H.W., Rudd S., Lemcke K., Mayer K.F.X., Quetier F.,
RA Salanoubat M.;
RL Submitted (NOV-1999) to the EMBL/GenBank/DDBJ databases.
(2)
RP SEQUENCE FROM N.A.
RA EU Arabidopsis sequencing project;
RL Submitted (APR-2000) to the EMBL/GenBank/DDBJ databases.
EMBL: AL132963; CAB87915.1; -.
DR PIR; T49283; T49283.
DR GO; GO:0005524; P:ATP binding; IEA.
DR GO; GO:0003754; P:chaperone activity; IEA.
DR GO; GO:0003677; F:DNA binding; IEA.
DR GO; GO:0004518; F:nuclease activity; IEA.
DR GO; GO:0000166; F:nucleotide binding; IEA.
DR GO; GO:0006289; P:nucleotide-excision repair; IEA.
DR InterPro; IPRO03593; AAA_ATPase.
DR InterPro; IPRO03959; AAA_ATPase centz.
DR InterPro; IPRO01270; Chaptain_clpA/B.
DR InterPro; IPRO04176; Clp_N.
DR InterPro; IPRO01943; UvrB/C.
DR Pfam; PF00004; AAA; 2.
DR Pfam; PF02861; Clp_N; 3.
DR Pfam; PF02151; UVR; 1.
DR PRINTS; PR00300; CLPPROTEASEA.
DR SMART; SM00382; AAA; 2.
DR PROSITE; PS00870; CLPAB_1; 1.
DR PROSITE; PS00871; CLPAB_2; 1.
DR PROSITE; PSS0151; UVR; 1.
KW ATP-binding.
SQ SEQUENCE 952 AA; 105771 MW; 295A0B61319E758C CRC64;

Query Match 1.5%; Score 8; DB 10; Length 952;
Best Local Similarity 100.0%; Pred.No. 1e+02;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;


```

OX NCBI_TaxID=32025;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 51449 / 3B1;
RX MEDLINE=22709201; PubMed=12810954;
RA Suerbaum S., Josenhans C., Sterzenbach T., Drescher B., Brandt P.,
RA Bell M., Droege M., Fartmann B., Fischer H.-P., Ge Z., Hoerster A.,
RA Holland R., Klein K., Koenig J., Macko L., Mendz G.L., Nyakatura G.,
RA Schauer D.B., Shen Z., Weber J., Froesch M., Fox J.G.;
RT "The complete genome sequence of the carcinogenic bacterium
RT Helicobacter hepaticus."
RL Proc. Natl. Acad. Sci. U.S.A. 100:7901-7906(2003).
DR EMBL; AB017145; AAP77222.1; -.
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 1021 AA; 112011 MW; 41ADF6BF1744F13C CRC64;

Query Match 1.5%; Score 8; DB 16; Length 1021;
Best Local Similarity 100.0%; Pred. No. 1.le+02;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 388 SSEVFTTP 395
Db 698 SSEVFTTF 705

```

```

RESULT 66
O76037 PRELIMINARY; PRT; 1045 AA.
ID ID O76037
AC AC O76037;
DT DT 01-NOV-1998 {TEMBLrel. 08, Created}
DT DT 01-NOV-1998 {TEMBLrel. 08, Last sequence update}
DT DT 01-JUN-2003 {TEMBLrel. 24, Last annotation update}
DE DE Mucin precursor (Fragment).
GN GN MUC4.
OS OS Homo sapiens (Human).
OC OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX OX NCBI_TaxId=9606;
RN RN [1]
RP RP SEQUENCE FROM N.A.
RC RC TISSUE=Colon mucosa;
RX RX MEDLINE=98285542; PubMed=9620877;
RA RA Nollet S., Montaux N., Maury J.P., Petitprez D., Degand P., Laine A.,
RA RA Porchet N., Aubert J.P.;
RT RT "Human Mucin Gene MUC4: Organization of its 5' Region and of its
RT RT Polymorphic Central Tandem Repeat Array.";
RL RL Biochem. J. 332:739-748(1998).
RN RN [2]
RP RP SEQUENCE FROM N.A.
RC RC TISSUE=Colon mucosa;
RX RX MEDLINE=20381033; PubMed=10920259;
RA RA Choudhury A., Montaux N., Winpenny J.P., Hollingsworth M.A.,
RA RA Aubert J.P., Batra S.K.;
RT RT "Human MUC4 mucin cDNA and its variants in pancreatic carcinoma.";
RT RT J. Biochem. 128:233-243(2000).
DR DR ENWL; AJ000281; CRA03985.1; -.
DR DR GO; GO:0005201; F:extracellular matrix structural constituent; NAS.
KW KW Signal.
FT FT 1 27 POTENTIAL.
FT FT CHAIN 28 >1045 MUCIN.
FT FT NON TER 1045 1045
SQ SQ SEQUENCE 1045 AA; 106737 MW; 26F12DF28E6D60A0 CRC64;

```

```

Query Match          1.5%; Score 8; DB 4; Length 1045;
Best Local Similarity 100.0%; Pred. No. 1.1e+02;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy      513 STRASQOT 520
        |||||
Db      783 STRASQOT 790

RESULT 67

```

```

DE sv12).
GN MUC4.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Pancreatic tumor;
RA Choudhury A., Moniaux N., Hollingsworth M.A., Aubert J.P., Batra S.K.;
RT "Human MUC4 mucin splice variants in pancreatic adenocarcinoma.";
RL Submitted (APR-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AJ400850; CAC14135.1; -.
DR EMBL; AJ400849; CAC14134.1; -.
SQ SEQUENCE 1157 AA; 118925 MW; F46F039320969A9C CRC64;

Query Match 1.5%; Score 8; DB 4; Length 1157;
Best Local Similarity 100.0%; Pred. No. 1.2e+02;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 513 STEASQGT 520
Db 783 STEASQGT 790

RESULT 70
Q9GZV6 PRELIMINARY; PRT; 1187 AA.
AC Q9GZV6;
DT 01-MAR-2001 (TrEMBLrel. 16, Created)
DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE MUC4 protein splice variant sv14 (MUC4 protein variant VII).
GN MUC4.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Pancreatic tumor;
RA Choudhury A., Moniaux N., Hollingsworth M.A., Aubert J.P., Batra S.K.;
RT "Human MUC4 mucin splice variants in pancreatic adenocarcinoma.";
RL Submitted (APR-2000) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=20381033; PubMed=10920259;
RA Choudhury A., Moniaux N., Wimpenny J.P., Hollingsworth M.A.,
RA Aubert J.P., Batra S.K.;
RT "Human MUC4 mucin cDNA and its variants in pancreatic carcinoma.";
RL J. Biochem. 128:233-243(2000).
DR EMBL; AJ400851; CAC14136.1; -.
DR EMBL; AJ400633; CAC10062.1; -.
DR GO; GO:0016020; C:membrane; NAS.
DR GO; GO:0005201; P:extracellular matrix structural constituent; NAS.
SQ SEQUENCE 1187 AA; 121925 MW; 62E4C93F86443B8F CRC64;

Query Match 1.5%; Score 8; DB 4; Length 1187;
Best Local Similarity 100.0%; Pred. No. 1.2e+02;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 513 STEASQGT 520
Db 783 STEASQGT 790

RESULT 71
Q9H487 PRELIMINARY; PRT; 1199 AA.
AC Q9H487;
DT 01-MAR-2001 (TrEMBLrel. 16, Created)
DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT 01-OCT-2002 (TrEMBLrel. 22, Last annotation update)

```

```

DE MUC4 protein splice variant sv15.
GN MUC4.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Pancreatic tumor;
RA Choudhury A., Moniaux N., Hollingsworth M.A., Aubert J.P., Batra S.K.;
RT "Human MUC4 mucin splice variants in pancreatic adenocarcinoma.";
RL Submitted (APR-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AJ400852; CAC14137.1; -.
SQ SEQUENCE 1199 AA; 123130 MW; 135C88F6977EF11B CRC64;

Query Match 1.5%; Score 8; DB 4; Length 1199;
Best Local Similarity 100.0%; Pred. No. 1.2e+02;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 513 STEASQGT 520
Db 783 STEASQGT 790

RESULT 72
Q9H486 PRELIMINARY; PRT; 1201 AA.
AC Q9H486;
DT 01-MAR-2001 (TrEMBLrel. 16, Created)
DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT 01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
DE MUC4 protein splice variant sv16.
GN MUC4.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Pancreatic tumor;
RA Choudhury A., Moniaux N., Hollingsworth M.A., Aubert J.P., Batra S.K.;
RT "Human MUC4 mucin splice variants in pancreatic adenocarcinoma.";
RL Submitted (APR-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AJ400853; CAC14138.1; -.
SQ SEQUENCE 1201 AA; 123504 MW; 8C355BEAB245C5DC CRC64;

Query Match 1.5%; Score 8; DB 4; Length 1201;
Best Local Similarity 100.0%; Pred. No. 1.2e+02;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 513 STEASQGT 520
Db 783 STEASQGT 790

RESULT 73
Q9H485 PRELIMINARY; PRT; 1214 AA.
AC Q9H485;
DT 01-MAR-2001 (TrEMBLrel. 16, Created)
DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT 01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
DE MUC4 protein splice variant sv17.
GN MUC4.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Pancreatic tumor;
RA Choudhury A., Moniaux N., Hollingsworth M.A., Aubert J.P., Batra S.K.;
RT "Human MUC4 mucin splice variants in pancreatic adenocarcinoma.";

```

RL Submitted (APR-2000) to the EMBL/GenBank/DBJ databases.
 DR EMBL: AJ400854; CAC14139.1; --
 SQ SEQUENCE 1214 AA; 125245 MW; B89D6A84AD4A056 CRC64;

Query Match 1.5%; Score 8; DB 4; Length 1214;
 Best Local Similarity 100.0%; Pred. No. 1.2e+02;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 513 STEASGQT 520
 |||||
 DB 783 STEASGQT 790

RESULT 74

Q9H484 PRELIMINARY; PRT; 1215 AA.

AC Q9H484; (TRENBLrel. 16, Created)
 DT 01-MAR-2001 (TRENBLrel. 16, Last sequence update)
 DT 01-MAR-2001 (TRENBLrel. 23, Last annotation update)
 DE MUC4 protein splice variant sv18.
 GN MUC4.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 CC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.

RC TISSUE=Pancratic tumor;
 RA Choudhury A., Moniaux N., Hollingsworth M.A., Aubert J.P., Batra S.K.;
 "Human MUC4 mucin splice variants in pancreatic adenocarcinoma";
 RT Submitted (APR-2000) to the EMBL/GenBank/DBJ databases.
 DR EMBL: AJ400855; CAC14140.1; --
 SQ SEQUENCE 1215 AA; 125435 MW; 2B950CD46945D60D CRC64;

Query Match 1.5%; Score 8; DB 4; Length 1215;
 Best Local Similarity 100.0%; Pred. No. 1.2e+02;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 513 STEASGQT 520
 |||||
 DB 783 STEASGQT 790

RESULT 75

Q9H483 PRELIMINARY; PRT; 1256 AA.

AC Q9H483; (TRENBLrel. 16, Created)
 DT 01-MAR-2001 (TRENBLrel. 16, Last sequence update)
 DT 01-MAR-2003 (TRENBLrel. 23, Last annotation update)
 DE MUC4 protein splice variant sv19.
 GN MUC4.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 CC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.

RC TISSUE=Pancratic tumor;
 RA Choudhury A., Moniaux N., Hollingsworth M.A., Aubert J.P., Batra S.K.;
 "Human MUC4 mucin splice variants in pancreatic adenocarcinoma";
 RT Submitted (APR-2000) to the EMBL/GenBank/DBJ databases.
 DR EMBL: AJ400856; CAC14141.1; --
 SQ SEQUENCE 1256 AA; 130269 MW; 5BA45CE53651AADC CRC64;

Query Match 1.5%; Score 8; DB 4; Length 1256;
 Best Local Similarity 100.0%; Pred. No. 1.3e+02;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 513 STEASGQT 520
 |||||
 DB 783 STEASGQT 790

RESULT 76

Q8PKZ7 PRELIMINARY; PRT; 1290 AA.

AC Q8PKZ7; (TRENBLrel. 22, Created)
 DT 01-OCT-2002 (TRENBLrel. 22, Last sequence update)
 DT 01-OCT-2002 (TRENBLrel. 22, Last annotation update)
 DE Hypothetical protein XAC2009.
 GN XAC2009.
 OS Xanthomonas axonopodis (pv. citri).
 CC Bacteria; Proteobacteria; Gammaproteobacteria; Xanthomonadales;
 CC Xanthomonadaceae; Xanthomonas.
 OX NCBI_TaxID=32829;
 RN [1]
 RP SEQUENCE FROM N.A.

RC STRAIN=306 / ATCC 13902 / XV 101;
 RX MEDLINE=22022145; PubMed=12024217;
 RA da Silva A.C.R., Ferro J.A., Reinach F.C., Farah C.S., Furlan L.R.,
 Quaggio R.B., Monteiro-Vitorello C.B., Van Sluys M.A., Almeida N.F.,
 Alves L.M.C., do Amaral A.M., Bertolini M.C., Camargo L.E.A.,
 Camarotte G., Cannavan F., Cardozo J., Chambergo F., Chapina L.P.,
 Cicarelli R.M.B., Coutinho L.L., Cursino-Santos J.R., El-Dorry H.,
 Faria J.B., Ferreira A.J.S., Ferreira R.C.C., Ferro M.I.T.,
 Fornighieri E.F., Franco M.C., Greggio C.C., Gruber A.,
 Katsuyama A.M., Kishi L.T., Leite R.P., Lemos E.G.M., Lemos M.V.P.,
 Locall E.C., Machado M.A., Madeira A.M.B.N., Martinez-Rossi N.M.,
 Martins E.C., Meidanis J., Menck C.F.M., Miyaki C.Y., Moon D.H.,
 Moreira L.M., Novo M.T.M., Okura V.K., Oliveira M.C., Oliveira V.R.,
 Pereira H.A., Rossi A., Sena J.A.D., Silva C., de Souza R.F.,
 Spinola L.A.P., Takita M.A., Tamura R.E., Teixeira E.C., Tezza R.I.D.,
 Trindade dos Santos M., Truffi D., Tsai S.M., White F.P.,
 RA Setubal J.C., Kitajima J.P.;
 RT "Comparison of the genomes of two Xanthomonas pathogens with differing
 host specificities";
 RL Nature 417:459-463(2002).
 DR EMBL: AB011836; AAM36871.1; --
 KW Hypothetical protein; Complete proteome.
 SQ SEQUENCE 1290 AA; 140955 MW; E9F16A206B21A01D CRC64;

Query Match 1.5%; Score 8; DB 16; Length 1290;
 Best Local Similarity 100.0%; Pred. No. 1.3e+02;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 321 VDAKIQQA 328
 |||||
 DB 313 VDAKIQQA 320

RESULT 77

Q9H482 PRELIMINARY; PRT; 1827 AA.

AC Q9H482; (TRENBLrel. 16, Created)
 DT 01-MAR-2001 (TRENBLrel. 16, Last sequence update)
 DT 01-JUN-2003 (TRENBLrel. 24, Last annotation update)
 DE MUC4 protein splice variant sv20.
 GN MUC4.

OS Homo sapiens (Human).
 CC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 CC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Pancratic tumor;
 RA Choudhury A., Moniaux N., Hollingsworth M.A., Aubert J.P., Batra S.K.;
 "Human MUC4 mucin splice variants in pancreatic adenocarcinoma";
 RT Submitted (APR-2000) to the EMBL/GenBank/DBJ databases.
 DR EMBL: AJ400857; CAC14142.1; --
 DR GO: 0016020; C-membrane; IEA.
 DR GO: 0005215; P-transporter activity; IEA.
 DR GO: 0007160; P-cell-matrix adhesion; IEA.

```

DR GO: GO:0006810; P:transport; IEA.
DR InterPro: IPR005533; AMOP.
DR InterPro: IPR003886; Nidogen ext.
DR InterPro: IPR005829; Sug transporter.
DR InterPro: IPR001846; VWF_D.
DR Pfam: PF03782; AMOP; 1.
DR Pfam: PF00094; vwd; 1.
DR SMART: SM00723; AMOP; 1.
DR SMART: SM00539; NIDO; 1.
DR SMART: SM00216; VWD; 1.
DR PROSITE: PS00217; SUGAR TRANSPORT 2; 1.
SQ SEQUENCE 1827 AA; 194100 MW; 63CE789B5C0F8741 CRC64;

Query Match      1.5%; Score 8; DB 4; Length 1827;
Best Local Similarity 100.0%; Pred.No.1.8e+02;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 513 STEASGQT 520
Db 783 STEASGQT 790

RESULT 78
Q92JPS PRELIMINARY; PRT; 1902 AA.
AC Q92JPS;
DT 01-DEC-2001 (TrEMBLrel. 19, Created)
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Cell surface antigen.
GN SCAL OR RC0019.
OS Rickettsia conorii.
OC Bacteria; Proteobacteria; Alphaproteobacteria; Rickettsiales;
OC Rickettsiaceae; Rickettsiae; Rickettsia.
OX NCBI_TaxID=781;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Malish 7;
RA MEDLINE=21442074; PubMed=11557893;
RA Ogata H., Audic S., Renesto-Audiffren P., Fournier P.-E., Barbe V.,
RA Samson D., Roux V., Cossart P., Weissbach J., Claverie J.-M.,
RA Raoult D.;
RT "Mechanisms of evolution in Rickettsia conorii and R. prowazekii.";
RL Science 293:2093-2098(2001).
DR ENBL: AE008570; AAL02557.1; -.
DR PIR: C97702; C97702.
DR GO: GO:0005509; F:calcium ion binding; IEA.
DR InterPro: IPR006315; Autotransporter.
DR InterPro: IPR005546; Autotransporter.
DR InterPro: IPR002048; EF-hand.
DR Pfam: PF03797; Autotransporter; 1.
DR TIGRFAMs: TIGR01414; autotrans_parl; 1.
DR PROSITE: PS00018; EF_HAND; 1.
DR Complete proteome.
KW Complete proteome.
SQ SEQUENCE 1902 AA; 212152 MW; F2D817457EF1912A CRC64;

Query Match      1.5%; Score 8; DB 16; Length 1902;
Best Local Similarity 100.0%; Pred.No.1.9e+02;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 405 KIKKLEKE 412
Db 356 KIKKLEKE 363

RESULT 79
Q9H4D8 PRELIMINARY; PRT; 2117 AA.
AC Q9H4D8;
DT 01-MAR-2001 (TrEMBLrel. 16, Created)
DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE Mucin 4, variant V3.

```

```

GN MUC4.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Pancratic tumor;
RX MEDLINE=20381033; PubMed=10920259;
RA Choudhury A., Moniaux N., Wippeny J.P., Hollingsworth M.A.,
RA Aubert J.P., Batra S.K.;
RT "Human MUC4 mucin cDNA and its variants in pancreatic carcinoma.";
RL J. Biochem. 128:233-243(2000).
DR EMBL: AJ277412; CACI0061.1; -.
DR GO: GO:0005578; C:extracellular matrix; NAS.
DR GO: GO:0030197; F:extracellular matrix constituent, lubricant. .; NAS.
DR InterPro: IPR005533; AMOP.
DR InterPro: IPR006209; EGF-like.
DR InterPro: IPR006210; IEGF.
DR InterPro: IPR005829; Sug transporter.
DR InterPro: IPR001846; VWF_D.
DR Pfam: PF03782; AMOP; 1.
DR Pfam: PF00008; EGF; 1.
DR Pfam: PF00094; vwd; 1.
DR SMART: SM00723; AMOP; 1.
DR SMART: SM00181; EGF; 1.
DR SMART: SM00216; VWD; 1.
DR PROSITE: PS00022; EGF 1; 1.
DR PROSITE: PS00217; SUGAR TRANSPORT 2; 1.
SQ SEQUENCE 2117 AA; 225404 MW; 086EB2A256F37BDE CRC64;

Query Match      1.5%; Score 8; DB 4; Length 2117;
Best Local Similarity 100.0%; Pred.No.2.1e+02;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 513 STEASGQT 520
Db 783 STEASGQT 790

RESULT 80
Q9H481 PRELIMINARY; PRT; 2167 AA.
AC Q9H481;
DT 01-MAR-2001 (TrEMBLrel. 16, Created)
DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE MUC4 protein splice variant sv21.
GN MUC4.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Pancratic tumor;
RA Choudhury A., Moniaux N., Hollingsworth M.A., Aubert J.P., Batra S.K.;
RT "Human MUC4 mucin splice variants in pancreatic adenocarcinoma.";
RL Submitted (APR-2000) to the EMBL/GenBank/DBJ databases.
DR ENBL: AJ400858; CAC14143.1; -.
DR GO: GO:0016020; C:membrane; IEA.
DR GO: GO:0005215; P:transporter activity; IEA.
DR GO: GO:0007160; P:cell-matrix adhesion; IEA.
DR GO: GO:0006810; P:transport; IEA.
DR InterPro: IPR005533; AMOP.
DR InterPro: IPR006209; EGF-like.
DR InterPro: IPR006210; IEGF.
DR InterPro: IPR003886; Nidogen_ext.
DR InterPro: IPR005829; Sug transporter.
DR InterPro: IPR001846; VWF_D.
DR Pfam: PF03782; AMOP; 1.
DR Pfam: PF00008; EGF; 1.
DR Pfam: PF00094; vwd; 1.

```

DR SMART; SM00723; AMOP; 1.
DR InterPro; IPR005629; EGF_like.
DR SMART; SM00181; EGF; 1.
DR SMART; SM00539; NIDO; 1.
DR SMART; SM00216; VMD; 1.
DR SMART; SM00022; EGF 1; 1.
DR PROSITE; PS00022; EGF 1; 1.
DR PROSITE; PS00217; SUGAR_TRANSPORT_2; 1.
DR PROSITE; PS00217; SUGAR_TRANSPORT_2; 1.
SEQUENCE 2167 AA; 231235 MW; 3AFPA34178C22AF7 CRC64;

Query Match
Best Local Similarity 100.0%; Score 8; DB 4; Length 2167;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 513 STEASGQT 520
Db 783 STEASGQT 790

RESULT 81
Q9NY09
ID Q9NY09 PRELIMINARY; PRT; 2169 AA.
AC Q9NY09
DT 01-OCT-2000 (TREMBLrel. 15, Created)
DT 01-OCT-2000 (TREMBLrel. 15, Last sequence update)
DT 01-JUN-2003 (TREMBLrel. 24, Last annotation update)
DE Mucin 4 (MUC4).
GN MUC4.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
NCBI_TaxID=9606;
RX [1]
SEQUENCE FROM N.A.
RP MEDLINE=20381033; PubMed=10920259;
RA Choudhury A., Moniaux N., Wimpenny J.P., Hollingsworth M.A.,
Albert J.P., Batra S.K.;
RT "Human MUC4 mucin cDNA and its variants in pancreatic carcinoma";
J. Biochem. 128:233-243(2000).
RN [2]
SEQUENCE FROM N.A.
RA Desseyn J.-L., Clavereau I., Laine A.;
RT "Cloning, chromosomal localization and characterization of the murine
mucin gene orthologous to human MUC4";
Eur. J. Biochem. 269:0-0(2002).
RL EMBL; AJ276359; CAB81773.1; -.
DR EMBL; AF522055; AAM66747.1; JOINED.
DR EMBL; AF522031; AAM66747.1; JOINED.
DR EMBL; AF522032; AAM66747.1; JOINED.
DR EMBL; AF522033; AAM66747.1; JOINED.
DR EMBL; AF522034; AAM66747.1; JOINED.
DR EMBL; AF522035; AAM66747.1; JOINED.
DR EMBL; AF522036; AAM66747.1; JOINED.
DR EMBL; AF522037; AAM66747.1; JOINED.
DR EMBL; AF522038; AAM66747.1; JOINED.
DR EMBL; AF522039; AAM66747.1; JOINED.
DR EMBL; AF522040; AAM66747.1; JOINED.
DR EMBL; AF522041; AAM66747.1; JOINED.
DR EMBL; AF522042; AAM66747.1; JOINED.
DR EMBL; AF522043; AAM66747.1; JOINED.
DR EMBL; AF522044; AAM66747.1; JOINED.
DR EMBL; AF522045; AAM66747.1; JOINED.
DR EMBL; AF522046; AAM66747.1; JOINED.
DR EMBL; AF522047; AAM66747.1; JOINED.
DR EMBL; AF522048; AAM66747.1; JOINED.
DR EMBL; AF522049; AAM66747.1; JOINED.
DR EMBL; AF522050; AAM66747.1; JOINED.
DR EMBL; AF522051; AAM66747.1; JOINED.
DR EMBL; AF522052; AAM66747.1; JOINED.
DR EMBL; AF522053; AAM66747.1; JOINED.
DR EMBL; AF522054; AAM66747.1; JOINED.
GO; GO:0016020; C:membrane; IEA.
GO; GO:0005215; F:transporter activity; IEA.
GO; GO:0006810; P:transport; IEA.
DR InterPro; IPR005533; AMOP.

DR SMART; SM00723; AMOP; 1.
DR InterPro; IPR005629; EGF_like.
DR SMART; SM00181; EGF; 1.
DR SMART; SM00539; NIDO; 1.
DR SMART; SM00216; VMD; 1.
DR SMART; SM00022; EGF 1; 1.
DR PROSITE; PS00022; EGF 1; 1.
DR PROSITE; PS00217; SUGAR_TRANSPORT_2; 1.
DR PROSITE; PS00217; SUGAR_TRANSPORT_2; 1.
SEQUENCE 2167 AA; 231235 MW; 3AFPA34178C22AF7 CRC64;

Query Match
Best Local Similarity 100.0%; Score 8; DB 4; Length 2167;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 513 STEASGQT 520
Db 783 STEASGQT 790

RESULT 82
Q26676
ID Q26676 PRELIMINARY; PRT; 23 AA.
AC Q26676
DT 01-NOV-1996 (TREMBLrel. 01, Created)
DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)
DT 01-JUN-2003 (TREMBLrel. 24, Last annotation update)
DE SPOXTAL protein (Fragment).
GN SPOXTAL.
OS Tetrahya aurantia (Orange puflball sponge).
OC Eukaryota; Metazoa; Porifera; Demospongiae; Tetractinomorpha;
OC Hadromerida; Tethyidae; Tethya.
NCBI_TaxID=34494;
RX [1]
SEQUENCE FROM N.A.
RP MEDLINE=95237605; PubMed=7721087;
RA Degnan B.M., Degnan S.M., Giusti A., Morse D.E.;
RT "A box/hom homeobox gene in sponges";
Gene 155:175-178(1995).
RL EMBL; X79265; CAA55852.1; -.
DR EMBL; AF522032; S45032.
FT NON_TER 1
FT NON_TER 23
FT NON_TER 23
SQ SEQUENCE 23 AA; 2684 MW; 02682802A01FF3C2 CRC64;

Query Match
Best Local Similarity 100.0%; Score 7; DB 5; Length 23;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 288 ELAERLK 294
Db 12 ELAERLK 18

RESULT 83
Q91K25
ID Q91K25 PRELIMINARY; PRT; 40 AA.
AC Q91K25
DT 01-DEC-2001 (TREMBLrel. 19, Created)
DT 01-DEC-2001 (TREMBLrel. 19, Last sequence update)
DT 01-JUN-2003 (TREMBLrel. 24, Last annotation update)
DE Genome polyprotein (Fragment).
OS Hepatitis C virus.
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC Replicavirus.
NCBI_TaxID=11103;
RX [1]
SEQUENCE FROM N.A.
RP STRAIN=H20;
RC Gimenez-Barcons M., Franco S.S., Suarez Y., Fornes X., Ampurdanes S.,
Paig-Basagoiti F., Sanchez-Fueyo A., Barrera J.M., Llovet J.M.,
Brux J.J., Sanchez-Tapias J.M., Rodes J., Saiz J.C.;
RT "High amino acid variability within the NS5A gene of hepatitis C virus
(HCV) is associated with hepatocellular carcinoma in patients with

RT HCV-1b related cirrhosis.";
 RL Hepatology 0:0-0(2001).
 DR EMBL; AF379247; AAK63464.1; -.
 DR GO; GO:0016021; C:integral to membrane; IEA.
 DR GO; GO:0019028; C:viral capsid; IEA.
 DR GO; GO:0019031; C:viral envelope; IEA.
 DR GO; GO:0005198; F:structural molecule activity; IEA.
 DR InterPro; IPR002531; HCV_NSI.
 DR Pfam; PF01560; HCV_NSI; 1.
 KW Coat protein; Envelope protein; Glycoprotein; Nonstructural protein;
 KW Polyprotein; Transmembrane.
 FT NON_TER 1 1
 FT NON_TER 40 40
 SQ SEQUENCE 40 AA; 4208 MW; A239A6F4CEB96829 CRC64;
 Query Match 1.3%; Score 7; DB 12; Length 40;
 Best Local Similarity 100.0%; Pred. No. 61;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 41 GAQARTA 47
 DB 18 GAQARTA 24
 RESULT 84
 Q8FHQ0 PRELIMINARY; PRT; 42 AA.
 AC Q8FHQ0;
 DT 01-MAR-2003 (TrEMBLrel. 23, Created)
 DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
 DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
 DE Hypothetical protein.
 GN C1813.
 OS Escherichia coli O6.
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
 OC Enterobacteriaceae; Escherichia.
 OX NCBI_TaxID=217992;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=O6:H1 / CFT073 / ATCC 700928;
 RX MEDLINE=2238234; PubMed=12471157;
 RA Welch R.A., Burland V., Plunkett G. III, Redford P., Roesch P.,
 RA Raako D., Buckles E.L., Liou S.-R., Boutin A., Hackett J., Stroud D.,
 RA Maynew G.F., Rose D.J., Zhou S., Schwartz D.C., Perna N.T.,
 RA Mobley H.L.T., Donnenberg M.S., Blattner F.R.;
 RT "Extensive mosaic structure revealed by the complete genome sequence
 of uropathogenic Escherichia coli.";
 RL Proc. Natl. Acad. Sci. U.S.A. 99:17020-17024(2002).
 DR EMBL; AE016760; AAN80277.1; -.
 KW Hypothetical protein; Complete proteome.
 SQ SEQUENCE 42 AA; 4824 MW; BEE47C8780D1A75 CRC64;
 Query Match 1.3%; Score 7; DB 16; Length 42;
 Best Local Similarity 100.0%; Pred. No. 64;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 424 NKALLEM 430
 DB 18 NKALLEM 24
 RESULT 85
 O72301 PRELIMINARY; PRT; 47 AA.
 AC O72301;
 DT 01-AUG-1998 (TrEMBLrel. 07, Created)
 DT 01-AUG-1998 (TrEMBLrel. 07, Last sequence update)
 DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
 DE Genome polyprotein (Fragment).
 OS Hepatitis C virus.
 OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
 OC Hepacivirus.
 OX NCBI_TaxID=11103;
 [1]
 RN SEQUENCE FROM N.A.
 RC STRAIN=anti-D Recipient 68;
 RX MEDLINE=98241727; PubMed=9573256;
 RA McAllister J., Casino C., Davidson P., Power J., Lawlor E., Yap P.L.,
 RA Simmonds P., Smith D.B.;
 RT "Long-term evolution of the hypervariable region of hepatitis C virus
 in a common-source-infected cohort.";
 RL J. Virol. 72:4893-4905(1998).
 DR EMBL; AF056879; AAC7375.1; -.
 DR GO; GO:0016021; C:integral to membrane; IEA.
 DR GO; GO:0019028; C:viral capsid; IEA.
 DR GO; GO:0019031; C:viral envelope; IEA.
 DR GO; GO:0005198; F:structural molecule activity; IEA.
 DR InterPro; IPR002519; HCV env.
 DR Pfam; PF01539; HCV env; 1.
 KW Coat protein; Envelope protein; Glycoprotein; Polyprotein;
 KW Transmembrane.
 FT NON_TER 1 1
 FT NON_TER 47 47
 SQ SEQUENCE 47 AA; 4854 MW; 7C74EB8227AC7A23 CRC64;
 Query Match 1.3%; Score 7; DB 12; Length 47;
 Best Local Similarity 100.0%; Pred. No. 71;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 41 GAQARTA 47
 DB 25 GAQARTA 31
 RESULT 86
 Q8XZX0 PRELIMINARY; PRT; 60 AA.
 AC Q8XZX0;
 DT 01-MAR-2002 (TrEMBLrel. 20, Created)
 DT 01-MAR-2002 (TrEMBLrel. 20, Last sequence update)
 DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
 DE Hypothetical protein ECs4994.
 GN ECs4994.
 OS Escherichia coli O157:H7.
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
 OC Enterobacteriaceae; Escherichia.
 OX NCBI_TaxID=83334;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=O157:H7 / RIMD 0509952;
 RX MEDLINE=21156331; PubMed=11258796;
 RA Hayashi T., Makino K., Ohnishi M., Kurokawa K., Ishii K., Yokoyama K.,
 RA Han C.-G., Ohtsubo E., Nakayama K., Murata T., Tanaka M., Tobe T.,
 RA Iida T., Takami H., Honda T., Sasakawa C., Ogasawara N., Yasunaga T.,
 RA Kuhara S., Shiba T., Hattori M., Shinagawa H.;
 RT "Complete genome sequence of enterohemorrhagic Escherichia coli
 O157:H7 and genomic comparison with a laboratory strain K-12.";
 RL DNA Res. 8:111-22(2001).
 DR EMBL; AF002567; BAB38417.1; -.
 DR PIR; B91253; B91253.
 KW Hypothetical protein.
 SQ SEQUENCE 60 AA; 6634 MW; BAC2C287091F20C0 CRC64;
 Query Match 1.3%; Score 7; DB 16; Length 60;
 Best Local Similarity 100.0%; Pred. No. 89;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 494 AQAPSSP 500
 DB 48 AQAPSSP 54
 RESULT 87
 Q8YVQ3 PRELIMINARY; PRT; 62 AA.
 ID Q8YVQ3
 AC Q8YVQ3;
 DT 01-MAR-2002 (TrEMBLrel. 20, Created)

RT HCV-1b related cirrhosis.";
 RL Hepatology 0:0-0(2001).
 DR EMBL; AF379247; AAK63464.1; -.
 DR GO; GO:0016021; C:integral to membrane; IEA.
 DR GO; GO:0019028; C:viral capsid; IEA.
 DR GO; GO:0019031; C:viral envelope; IEA.
 DR GO; GO:0005198; F:structural molecule activity; IEA.
 DR InterPro; IPR002531; HCV_NSI.
 DR Pfam; PF01560; HCV_NSI; 1.
 KW Coat protein; Envelope protein; Glycoprotein; Nonstructural protein;
 KW Polyprotein; Transmembrane.
 FT NON_TER 1 1
 FT NON_TER 40 40
 SQ SEQUENCE 40 AA; 4208 MW; A239A6F4CEB96829 CRC64;
 Query Match 1.3%; Score 7; DB 12; Length 40;
 Best Local Similarity 100.0%; Pred. No. 61;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 41 GAQARTA 47
 DB 18 GAQARTA 24
 RESULT 84
 Q8FHQ0 PRELIMINARY; PRT; 42 AA.
 AC Q8FHQ0;
 DT 01-MAR-2003 (TrEMBLrel. 23, Created)
 DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
 DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
 DE Hypothetical protein.
 GN C1813.
 OS Escherichia coli O6.
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
 OC Enterobacteriaceae; Escherichia.
 OX NCBI_TaxID=217992;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=O6:H1 / CFT073 / ATCC 700928;
 RX MEDLINE=2238234; PubMed=12471157;
 RA Welch R.A., Burland V., Plunkett G. III, Redford P., Roesch P.,
 RA Raako D., Buckles E.L., Liou S.-R., Boutin A., Hackett J., Stroud D.,
 RA Maynew G.F., Rose D.J., Zhou S., Schwartz D.C., Perna N.T.,
 RA Mobley H.L.T., Donnenberg M.S., Blattner F.R.;
 RT "Extensive mosaic structure revealed by the complete genome sequence
 of uropathogenic Escherichia coli.";
 RL Proc. Natl. Acad. Sci. U.S.A. 99:17020-17024(2002).
 DR EMBL; AE016760; AAN80277.1; -.
 KW Hypothetical protein; Complete proteome.
 SQ SEQUENCE 42 AA; 4824 MW; BEE47C8780D1A75 CRC64;
 Query Match 1.3%; Score 7; DB 16; Length 42;
 Best Local Similarity 100.0%; Pred. No. 64;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 424 NKALLEM 430
 DB 18 NKALLEM 24
 RESULT 85
 O72301 PRELIMINARY; PRT; 47 AA.
 AC O72301;
 DT 01-AUG-1998 (TrEMBLrel. 07, Created)
 DT 01-AUG-1998 (TrEMBLrel. 07, Last sequence update)
 DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
 DE Genome polyprotein (Fragment).
 OS Hepatitis C virus.
 OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
 OC Hepacivirus.
 OX NCBI_TaxID=11103;

01-MAR-2002 (TRENBLrel. 20, Last sequence update)
01-JUN-2003 (TRENBLrel. 24, Last annotation update)
Hypothetical protein ASL0793.
OS Anabaena sp. (strain PCC 7120).
OC Bacteria; Cyanobacteria; Nostocales; Nostocaceae; Nostoc.
OX NCBI_TaxID=103690;
RN [1]
SEQUENCE FROM N.A.
RX MEDLINE=21595285; PubMed=11759840;
RA Kaneko T., Nakamura Y., Wolk C.P., Kuritz T., Sasamoto S.,
RA Watanabe A., Iriiguchi M., Ishikawa A., Kawashima K., Kimura T.,
RA Kishida Y., Kohara M., Matsumoto M., Matsuno A., Muraki A.,
RA Nakazaki N., Shimo S., Sugimoto M., Takazawa M., Yamada M.,
RA Yasuda M., Tabata S.;
RT "Complete genomic sequence of the filamentous nitrogen-fixing
cyanobacterium Anabaena sp. strain PCC 7120.";
RL DNA Res. 8:205-213 (2001).
DR EMBL; AP003583; BAB7250.1; --
DR PIR; AG1905; AG1905.
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 62 AA; 7067 MW; 0229460C38C80422 CRC64;

Query Match 1.3%; Score 7; DB 16; Length 62;
Best Local Similarity 100.0%; Pred. No. 91;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 471 DLSAGGQ 477
|||
DB 21 DLSAGGQ 27

RESULT 88
Q72305 PRELIMINARY; PRT; 69 AA.
AC Q72305;
DT 01-AUG-1998 (TRENBLrel. 07, Created)
DT 01-AUG-1998 (TRENBLrel. 07, Last sequence update)
DT 01-JUN-2003 (TRENBLrel. 24, Last annotation update)
DE Genome polyprotein (Fragment).
OS Hepatitis C virus.
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC Hepacivirus.
OX NCBI_TaxID=11103;
RN [1]
SEQUENCE FROM N.A.
RC STRAIN=anti-D Recipient 68;
RX MEDLINE=98241727; PubMed=9573256;
RA McAllister J., Casino C., Davidson F., Power J., Lawlor E., Yap P.L.,
RA Simmonds P., Smith D.B.;
RT "Long-term evolution of the hypervariable region of hepatitis C virus
in a common-source-infected cohort.";
RL J. Virol. 72:4893-4905 (1998).
DR EMBL; AF056883; AAC17379.1; --
DR GO; GO:0016021; C:integral to membrane; IEA.
DR GO; GO:0019028; C:viral capsid; IEA.
DR GO; GO:0019031; C:viral envelope; IEA.
DR GO; GO:0005198; F:structural molecule activity; IEA.
DR InterPro; IPR002531; HCV_NSI.
DR Pfam; PF01560; HCV_NSI; 1.
DR ProDom; PD186062; HCV_NSI; 1.
KW Coat protein; Envelope protein; Glycoprotein; Nonstructural protein;
KW Polyprotein; Transmembrane.
FT NON_TER 1
FT NON_TER 69
SQ SEQUENCE 69 AA; 7357 MW; F257BCCF7E4BFFAE CRC64;

Query Match 1.3%; Score 7; DB 12; Length 69;
Best Local Similarity 100.0%; Pred. No. 1e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 41 GAQARTA 47
|||
|||

Db 12 GAQARTA 18

RESULT 89
Q8VCPI PRELIMINARY; PRT; 70 AA.
AC Q8VCPI;
DT 01-MAR-2002 (TRENBLrel. 20, Created)
DT 01-MAR-2002 (TRENBLrel. 20, Last sequence update)
DT 01-OCT-2003 (TRENBLrel. 25, Last annotation update)
DE Hypothetical protein.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
SEQUENCE FROM N.A.
RC TISSUE=Breast tumor;
RA Strausberg R.;
RL Submitted (DEC-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; BC019472; AAH19472.1; --
KW Hypothetical protein.
SQ SEQUENCE 70 AA; 7944 MW; BCE7463EA6BA6CAB CRC64;

Query Match 1.3%; Score 7; DB 11; Length 70;
Best Local Similarity 100.0%; Pred. No. 1e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 236 NRSLSKEE 242
|||
DB 43 NRSLSKEE 49

RESULT 90
Q9WLK7 PRELIMINARY; PRT; 74 AA.
AC Q9WLK7;
DT 01-NOV-1999 (TRENBLrel. 12, Created)
DT 01-NOV-1999 (TRENBLrel. 12, Last sequence update)
DT 01-JUN-2003 (TRENBLrel. 24, Last annotation update)
DE Genome polyprotein (Fragment).
OS Hepatitis C virus.
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC Hepacivirus.
OX NCBI_TaxID=11103;
RN [1]
SEQUENCE FROM N.A.
RC STRAIN=anti-D Recipient 68;
RX MEDLINE=98241727; PubMed=9573256;
RA McAllister J., Casino C., Davidson F., Power J., Lawlor E., Yap P.L.,
RA Simmonds P., Smith D.B.;
RT "Long-term evolution of the hypervariable region of hepatitis C virus
in a common-source-infected cohort.";
RL J. Virol. 72:4893-4905 (1998).
DR EMBL; AF056885; AAC17381.1; --
DR GO; GO:0016021; C:integral to membrane; IEA.
DR GO; GO:0019028; C:viral capsid; IEA.
DR GO; GO:0019031; C:viral envelope; IEA.
DR GO; GO:0005198; F:structural molecule activity; IEA.
DR InterPro; IPR002531; HCV_NSI.
DR Pfam; PF01560; HCV_NSI; 1.
DR ProDom; PD186062; HCV_NSI; 1.
KW Coat protein; Envelope protein; Glycoprotein; Nonstructural protein;
KW Polyprotein; Transmembrane.
FT NON_TER 1
FT NON_TER 74
SQ SEQUENCE 74 AA; 7777 MW; 89933164D292B86F CRC64;

Query Match 1.3%; Score 7; DB 12; Length 74;
Best Local Similarity 100.0%; Pred. No. 1.1e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 41 GAQARTA 47
|||
|||

DB 11 GAOQRTA 17

RESULT 91

Q9W916 PRELIMINARY; PRT; 75 AA.

ID Q9W916

AC Q9W916

DT 01-NOV-1999 (T-EMBLrel. 12, Created)

DT 01-NOV-1999 (T-EMBLrel. 12, Last sequence update)

DT 01-JUN-2003 (T-EMBLrel. 24, Last annotation update)

DE Genome polyprotein (Fragment)

OS Hepatitis C virus.

OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;

OC Hepacivirus.

OX NCBI_TaxID=11103;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=ANTI-D RECIPIENT 68;

RX MEDLINE=98241727; PubMed=9573256;

RA McAllister J., Casino C., Davidson F., Power J., Lawlor E., Yap P.L.,

RA Simmonds P., Smith D.B.;

RT "Long-term evolution of the hypervariable region of hepatitis C virus

RL J. Virol. 72:4893-4905(1998).

DR EMBL; AF056889; AAC17385.1; -

DR EMBL; AF056884; AAC17380.1; -

DR EMBL; AF056886; AAC17382.1; -

DR EMBL; AF056888; AAC17384.1; -

DR GO; GO:0016021; C:integral to membrane; IEA.

DR GO; GO:0019028; C:viral capsid; IEA.

DR GO; GO:0019031; C:viral envelope; IEA.

DR GO; GO:0005198; F:structural molecule activity; IEA.

DR InterPro; IPR002531; HCV NS1.

DR Pfam; PF01560; HCV NS1; 1.

DR ProDom; PD186062; HCV NS1; 1.

KW Coat protein; Envelope protein; Glycoprotein; Nonstructural protein;

KW Polyprotein; Transmembrane.

FT NON_TER 1

FT NON_TER 75

FT SEQUENCE 75 AA; 7848 MW; 89936131D7C2ED6A CRC64;

Query Match 1.3%; Score 7; DB 12; Length 75;

Best Local Similarity 100.0%; Pred. No. 1.1e+02;

Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 41 GAOQRTA 47

DB 12 GAOQRTA 18

RESULT 92

Q928F1 PRELIMINARY; PRT; 75 AA.

AC Q928F1

DT 01-DEC-2001 (T-EMBLrel. 19, Created)

DT 01-DEC-2001 (T-EMBLrel. 19, Last sequence update)

DT 01-JUN-2003 (T-EMBLrel. 24, Last annotation update)

DE Hypothetical protein lin2584.

GN LIN2584.

OS Listeria innocua.

OC Bacteria; Firmicutes; Bacillales; Listeriaceae; Listeria.

OX NCBI_TaxID=1642;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=CLIP 11262 / Serovar 6a;

RX MEDLINE=21537279; PubMed=11679669;

RA Glaser P., Frangul L., Buchrieser C., Rusniok C., Amend A.,

RA Baquero F., Berche P., Bloeker H., Brandt P., Chakraborty T.,

RA Charbit A., Chetouani P., Couve E., de Daruvar A., Dehoux P.,

RA Domann E., Dominguez-Bernal G., Duchaud E., Durant L., Dussurget O.,

RA Entian K.-D., Fiebi H., Garcia-del Portillo F., Garrido P.,

RA Gautier L., Goebel W., Gomez-Lopez N., Hain T., Hauf J., Jackson D.,

Query Match 1.3%; Score 7; DB 12; Length 75;

Best Local Similarity 100.0%; Pred. No. 1.1e+02;

Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 41 GAOQRTA 47

DB 12 GAOQRTA 18

RESULT 93

Q92982 PRELIMINARY; PRT; 76 AA.

AC Q92982

DT 01-NOV-1998 (T-EMBLrel. 08, Created)

DT 01-NOV-1998 (T-EMBLrel. 08, Last sequence update)

DT 01-JUN-2003 (T-EMBLrel. 24, Last annotation update)

DE Genome polyprotein (Fragment)

OS Hepatitis C virus.

OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;

OC Hepacivirus.

OX NCBI_TaxID=11103;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=anti-D Recipient 68;

RX MEDLINE=98241727; PubMed=9573256;

RA McAllister J., Casino C., Davidson F., Power J., Lawlor E., Yap P.L.,

RA Simmonds P., Smith D.B.;

RT "Long-term evolution of the hypervariable region of hepatitis C virus

RL J. Virol. 72:4893-4905(1998).

DR EMBL; AF056887; AAC17383.1; -

DR GO; GO:0016021; C:integral to membrane; IEA.

DR GO; GO:0019028; C:viral capsid; IEA.

DR GO; GO:0019031; C:viral envelope; IEA.

DR GO; GO:0005198; F:structural molecule activity; IEA.

DR InterPro; IPR002531; HCV NS1.

DR Pfam; PF01560; HCV NS1; 1.

DR ProDom; PD186062; HCV NS1; 1.

KW Coat protein; Envelope protein; Glycoprotein; Nonstructural protein;

KW Polyprotein; Transmembrane.

FT NON_TER 1

FT NON_TER 76

FT SEQUENCE 76 AA; 7995 MW; 2BB361675F1CBBE2 CRC64;

Query Match 1.3%; Score 7; DB 12; Length 76;

Best Local Similarity 100.0%; Pred. No. 1.1e+02;

Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 41 GAOQRTA 47

DB 13 GAOQRTA 19

RESULT 94

Q9RSW4 PRELIMINARY; PRT; 79 AA.

ID Q9RSW4

AC Q9RSW4

DT 01-MAY-2000 (T-EMBLrel. 13, Created)

DT 01-MAY-2000 (T-EMBLrel. 13, Last sequence update)

DT 01-JUN-2003 (T-EMBLrel. 24, Last annotation update)

Query Match 1.3%; Score 7; DB 12; Length 76;

Best Local Similarity 100.0%; Pred. No. 1.1e+02;

Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 41 GAOQRTA 47

DB 13 GAOQRTA 19


```

DE Hypothetical protein DR2007.
GN DR2007.
OS Deinococcus radiodurans.
OC Bacteria; Deinococcus-Thermus; Deinococci; Deinococcales;
OC Deinococcaceae; Deinococcus.
OX NCBI_TaxID=1299;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=R1 / ATCC 13939 / DSM 20539 / NCIB 9279;
RX MEDLINE=20036896; PubMed=10567266;
RA White O., Eisen J.A., Heidelberg J.F., Hickey E.K., Peterson J.D.,
RA Dodson R.J., Haft D.H., Gwinn M.L., Nelson W.C., Richardson D.L.,
RA Moffat K.S., Qin H., Jiang L., Pamphile W., Crosby M., Shen M.,
RA Vamathevan J.J., Lam P., McDonald L., Utterback T., Zalewski C.,
RA Makarova K.S., Aravind L., Daly M.J., Minton K.W., Fleischmann R.D.,
RA Ketchum K.A., Nelson K.E., Salzberg S., Smith H.O., Venter J.C.,
RA Fraser C.M.;
RT "Genome sequence of the radioresistant bacterium Deinococcus
RT radiodurans R1.";
RL Science 286:1571-1577(1999).
DR EMBL; AB002038; AAF11565.1; -.
DR PIR; D75326; D75326.
DR TIGR; DR2007; -.
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 79 AA; 8688 MW; 0E7D1P4162591CSB CRC64;

Query Match 1.3%; Score 7; DB 16; Length 79;
Best Local Similarity 100.0%; Pred. No. 1.1e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 58 ELSRQLE 64
DB 12 ELSRQLE 18

RESULT 95
Q83L13 PRELIMINARY; PRT; 81 AA.
AC Q83L13;
DT 01-JUN-2003 (TrEMBLrel. 24, Created)
DT 01-JUN-2003 (TrEMBLrel. 24, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Orf, partial conserved hypothetical protein.
GN Y121.G5 OR SF1580.
OS Shigella flexneri.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Shigella.
OX NCBI_TaxID=623;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=301 / Serotype 2a;
RX MEDLINE=2272406; PubMed=12384590;
RA Jin Q., Yuan Z., Xu J., Wang Y., Shen Y., Lu W., Wang J., Liu H.,
RA Yang J., Yang F., Zhang X., Zhang J., Yang G., Wu H., Qu D., Dong J.,
RA Sun L., Xue Y., Zhao A., Gao Y., Zhu J., Kan B., Ding K., Chen S.,
RA Cheng H., Yao Z., He B., Chen R., Ma D., Qiang B., Wen Y., Hou Y.,
RA Yu J.;
RT "Genome sequence of Shigella flexneri 2a: insights into pathogenicity
RT through comparison with genomes of Escherichia coli K12 and O157.";
RL Nucleic Acids Res. 30:4432-4441(2002).
DR EMBL; AS015179; AN43168.1; -.
DR InterPro; IPR002514; Transposase 8.
DR Pfam; PF01527; Transposase 8; 1.
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 81 AA; 8937 MW; AEEC05F025593C82 CRC64;

Query Match 1.3%; Score 7; DB 16; Length 81;
Best Local Similarity 100.0%; Pred. No. 1.2e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 346 LKKEAVE 352
DB 56 LKKEAVE 62

DE Hypothetical protein DR2007.
GN DR2007.
OS Deinococcus radiodurans.
OC Bacteria; Deinococcus-Thermus; Deinococci; Deinococcales;
OC Deinococcaceae; Deinococcus.
OX NCBI_TaxID=1299;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=R1 / ATCC 13939 / DSM 20539 / NCIB 9279;
RX MEDLINE=20036896; PubMed=10567266;
RA White O., Eisen J.A., Heidelberg J.F., Hickey E.K., Peterson J.D.,
RA Dodson R.J., Haft D.H., Gwinn M.L., Nelson W.C., Richardson D.L.,
RA Moffat K.S., Qin H., Jiang L., Pamphile W., Crosby M., Shen M.,
RA Vamathevan J.J., Lam P., McDonald L., Utterback T., Zalewski C.,
RA Makarova K.S., Aravind L., Daly M.J., Minton K.W., Fleischmann R.D.,
RA Ketchum K.A., Nelson K.E., Salzberg S., Smith H.O., Venter J.C.,
RA Fraser C.M.;
RT "Genome sequence of the radioresistant bacterium Deinococcus
RT radiodurans R1.";
RL Science 286:1571-1577(1999).
DR EMBL; AB002038; AAF11565.1; -.
DR PIR; D75326; D75326.
DR TIGR; DR2007; -.
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 79 AA; 8688 MW; 0E7D1P4162591CSB CRC64;

Query Match 1.3%; Score 7; DB 16; Length 79;
Best Local Similarity 100.0%; Pred. No. 1.1e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 58 ELSRQLE 64
DB 12 ELSRQLE 18

RESULT 96
Q87W81 PRELIMINARY; PRT; 82 AA.
AC Q87W81;
DT 01-JUN-2003 (TrEMBLrel. 24, Created)
DT 01-JUN-2003 (TrEMBLrel. 24, Last sequence update)
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE Hypothetical protein.
GN PSPT04673.
OS Pseudomonas syringae (pv. tomato).
OC Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales;
OC Pseudomonadaceae; Pseudomonas.
OX NCBI_TaxID=323;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=DC3000;
RA Buell R., Joardar V., Khouri H., Fedorova N., Tran B., Russell D.,
RA Bezy K., Utterback T., Van Aken S., Feldblyum T., Gwinn M.,
RA Dodson R., DeBoy R., Durkin A., Kolonay J., Madupu R., Daugherty S.,
RA Brinkac L., Beanan M., Haft D., Selengut J., Nelson W., Davidsen T.,
RA White O., Fraser C., Collier A.;
RT "Complete sequence of Pseudomonas syringae.";
RL Submitted (MAR-2003) to the EMBL/GenBank/DBJ databases.
DR EMBL; AE016872; AA058119.1; -.
DR TIGR; PSPT04673; -.
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 82 AA; 8888 MW; 3826079B39FAD75C CRC64;

Query Match 1.3%; Score 7; DB 16; Length 82;
Best Local Similarity 100.0%; Pred. No. 1.2e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 384 TLSKSS 390
DB 35 TLSKSS 41

RESULT 97
Q86L13 PRELIMINARY; PRT; 84 AA.
AC Q86L13;
DT 01-AUG-1998 (TrEMBLrel. 07, Created)
DT 01-AUG-1998 (TrEMBLrel. 07, Last sequence update)
DT 01-NOV-1998 (TrEMBLrel. 08, Last annotation update)
DE BchZ (Fragment).
GN BchZ.
OS Acidiphilium rubrum.
OC Bacteria; Proteobacteria; Alphaproteobacteria; Rhodospirillales;
OC Acetobacteraceae; Acidiphilium.
OX NCBI_TaxID=526;
RN [1]
RP SEQUENCE FROM N.A.
RC MEDLINE=98096867; PubMed=9435141;
RX Nagashima K.V., Matsuura K., Wakao N., Hiraishi A., Shimada K.;
RT "Nucleotide sequences of genes coding for photosynthetic reaction
RT centers and light-harvesting proteins of Acidiphilium rubrum and
RT related aerobic acidophilic bacteria.";
RL Plant Cell Physiol. 38:1249-1258(1997).
DR EMBL; AB005218; BAA25547.1; -.
FT NON_TER 1
SQ SEQUENCE 84 AA; 9230 MW; A404E530138BPF61F CRC64;

Query Match 1.3%; Score 7; DB 2; Length 84;
Best Local Similarity 100.0%; Pred. No. 1.2e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 499 SPRVTEA 505
DB 66 SPRVTEA 72

```

```

RESULT 98
Q92JR7 PRELIMINARY; PRT; 84 AA.
AC Q92JR7
DT 01-DEC-2001 (TrEMBLrel. 19, Created)
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT 01-MAR-2002 (TrEMBLrel. 20, Last annotation update)
DE Hypothetical protein R03326.
GN R03326 OR SMC04092.
OS Rhizobium meliloti (Sinorhizobium meliloti).
OC Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;
OC Rhizobiaceae; Sinorhizobium/Ensifer group; Sinorhizobium.
OX NCBI_TaxID=382;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=1021;
RX MEDLINE=21396507; PubMed=11481430;
RA Capella D., Barloy-Hubler F., Gouzy J., Bothe G., Ampe P., Batut J.,
RA Boistard P., Becker A., Boutry M., Cadieu E., Dreano S., Gloux S.,
RA Godtje T., Goffeau A., Kahn D., Kiss E., Lelaure V., Masuy D.,
RA Pohl T., Portetelle D., Puehler A., Purnelle B., Ramsberger U.,
RA Renard C., Thebaud P., Vandenbol M., Weidner S., Calbert F.;
RT Analysis of the chromosome sequence of the legume symbiont
RT Sinorhizobium meliloti strain 1021.;
RL Proc. Natl. Acad. Sci. U.S.A. 98:9877-9882(2001).
DR EMBL; AL591793; CAC47905.1; -.
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 84 AA; 9237 MW; D4697D23X725997E CRC64;

Query Match 1.3%; Score 7; DB 16; Length 84;
Best Local Similarity 100.0%; Pred. No. 1.2e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 332 LXAER 338
DB 42 LXAER 48
|||||

RESULT 99
O66135 PRELIMINARY; PRT; 85 AA.
AC O66135
DT 01-AUG-1998 (TrEMBLrel. 07, Created)
DT 01-AUG-1998 (TrEMBLrel. 07, Last sequence update)
DT 01-NOV-1998 (TrEMBLrel. 08, Last annotation update)
DE BchZ (Fragment).
GN BCHZ.
OS Acidiphilium angustum.
OC Bacteria; Proteobacteria; Alphaproteobacteria; Rhodospirillales;
OC Acetobacteraceae; Acidiphilium.
OX NCBI_TaxID=523;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=98096867; PubMed=9435141;
RA Nagashima K.V., Matsura K., Wakao N., Hiraishi A., Shimada K.;
RT "Nucleotide sequences of genes coding for photosynthetic reaction
RT centers and light-harvesting proteins of Acidiphilium rubrum and
RT related aerobic acidophilic bacteria.";
RL Plant Cell Physiol. 38:1249-1258(1997).
DR EMBL; AB005219; BAA25554.1; -.
FT NON TER 1
RP SEQUENCE 85 AA; 9343 MW; B84C693D6C9EDC20 CRC64;

Query Match 1.3%; Score 7; DB 2; Length 85;
Best Local Similarity 100.0%; Pred. No. 1.2e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 499 SPRVTEA 505
DB 67 SPRVTEA 73
|||||

RESULT 100

```

```

Q8L5Q4 PRELIMINARY; PRT; 87 AA.
ID Q8L5Q4
AC Q8L5Q4
DT 01-OCT-2002 (TrEMBLrel. 22, Created)
DT 01-OCT-2002 (TrEMBLrel. 22, Last sequence update)
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE Putative adenosine kinase (EC 2.7.1.20) (Fragment).
GN ADK.
OS Cicer arietinum (Chickpea) (Garbanzo).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids;
OC eurosids I; Fabales; Fabaceae; Papilionoideae; Cicereae; Cicer.
OX NCBI_TaxID=3827;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=cv. Castellana; TISSUE=Epicotyl;
RA Dopico B., Labrador E.;
RT "An adenosine kinase is expressed in chickpea epicotyls.";
RT Submitted (MAY-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; AJ487468; CAD31841.1; -.
DR GO; GO:0004001; P:adenosine kinase activity; IEA.
DR GO; GO:0016301; P:kinase activity; IEA.
DR GO; GO:0016740; P:transferase activity; IEA.
DR InterPro; IPR002173; PFKB.
DR Pfam; PF00294; PfkB; 1.
DR PROSITE; PS00584; PPKB_KINASES_2; 1.
KW Kinase; Transferase.
FT NON TER 1
SQ SEQUENCE 87 AA; 9419 MW; 5703AA976B701A44 CRC64;

Query Match 1.3%; Score 7; DB 10; Length 87;
Best Local Similarity 100.0%; Pred. No. 1.2e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 202 SOLVQEK 208
DB 50 SOLVQEK 56
|||||

Search completed: June 7, 2004, 14:46:32
Job time : 50 secs

```

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: June 7, 2004, 12:38:45 ; Search time 139 Seconds
(without alignments)
1077.339 Million cell updates/sec

Title: US-10-023-529-8

Perfect score: 2702

Sequence: 1 KSSPCQBPAGPEGAQERDSQ.....APSTEASQGTQBPPTSARA 530

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1586107 seqs, 282547505 residues

Total number of hits satisfying chosen parameters: 1586107

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 150 summaries

Database : A_Geneseq_29Jan04.*

1: Geneseq1980s.*
2: Geneseq1990s.*
3: Geneseq2000s.*
4: Geneseq2001s.*
5: Geneseq2002s.*
6: Geneseq2003as.*
7: Geneseq2003bs.*
8: Geneseq2004s.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2702	100.0	530	2 AAW49042	Aaw49042 Human low
2	2702	100.0	530	4 AAB82804	Aab82804 Human low
3	2693	99.7	546	4 AAB82808	Aab82808 Human low
4	2693	99.7	639	6 ABR41210	AbR41210 Human DIT
5	2693	99.7	639	6 ABR41186	AbR41186 Human DIT
6	2465.5	91.2	557	2 AAW49039	Aaw49039 Rabbit lo
7	2465.5	91.2	557	4 AAB82801	Aab82801 Rabbit lo
8	2338.5	86.5	1749	4 ABG00839	Abg00839 Novel hum
9	2334	86.0	510	4 ABB11764	Abb11764 Human LDL
10	2324	86.0	510	4 AAM79741	Aam79741 Human pro
11	1801	66.7	356	6 ABO07223	AbO07223 Human p53
12	1754.5	64.9	358	4 AAM78757	Aam78757 Human pro
13	1283.5	47.5	528	4 AAB93594	Aab93594 Human pro
14	1265.5	46.8	395	6 ABO07222	AbO07222 Human p53
15	1260.5	46.7	684	6 ABR43284	AbR43284 Human neu
16	1233	45.6	676	2 AAV17863	Aav17863 Neurite e
17	1038.5	38.4	386	2 AAW37883	Aaw37883 BRCA1 mod
18	1038.5	38.4	386	2 AAY30151	Aay30151 Anitro aci
19	815	30.2	204	4 AAB63259	Aab63259 Human bre
20	810	30.0	475	4 AAO08301	Aao08301 Human pol
21	810	30.0	475	7 ADC33216	AdC33216 Human nov
22	659	24.4	416	7 ADC31613	AdC31613 Human nov
23	642.5	23.8	410	4 ABG02486	Abg02486 Novel hum
24	597.5	22.1	515	4 ABB61894	Abb61894 Drosophil
25	550	20.4	841	4 ABG02487	Abg02487 Novel hum

26	322	11.9	102	4	ABG02488	Abg02488 Novel hum
27	272	10.1	2274	4	ABB58657	Abb58657 Drosophil
28	265.5	9.9	954	4	AAU14615	Aau14615 Novel bon
29	265.5	9.9	1455	7	ADE80783	Ade80783 Microsate
30	257.5	9.5	860	7	ADC38517	Adc38517 Human AML
31	256	9.5	1752	2	AAU07031	Aau07031 Breast ca
32	256	9.5	2871	6	ABU07402	Abu07402 Protein d
33	256	9.5	2871	7	ADC35075	Adc35075 Human bre
34	256	9.5	2918	4	ABG27218	Abg27218 Novel hum
35	254.5	9.4	931	4	AAW79504	Aam79504 Human pro
36	254.5	9.4	990	4	AAW78520	Aam78520 Human pro
37	253.5	9.4	1881	5	ABP73809	Abp73809 Candida a
38	253	9.4	1206	7	ADD46507	Add46507 Rat Prote
39	253	9.4	1206	7	ADD48875	Add48875 Rat Prote
40	252.5	9.3	880	4	AAB96332	Aab96332 Putative
41	251.5	9.3	1690	4	ABB61144	Abb61144 Drosophil
42	251.5	9.3	1690	4	ABB61173	Abb61173 Drosophil
43	249	9.2	1392	2	AAV06999	Aav06999 Restin pr
44	248.5	9.2	786	2	AAV55937	Aay55937 Human SUL
45	248.5	9.2	1001	2	AAV55942	Aay55942 Human/Mur
46	248.5	9.2	1001	5	ABB97326	Abb97326 Novel hum
47	248.5	9.2	1001	7	ADE55356	Ade55356 Human pro
48	248.5	9.2	1005	6	ABR47509	AbR47509 Breast ca
49	248.5	9.2	1005	6	AAO30953	Aao30953 Human TAO
50	247	9.1	85	5	ABP42281	Abp42281 Human ova
51	247	9.1	1001	2	AAV48896	Aay48896 Rat TA01
52	247	9.1	1001	7	ADE55355	Ade55355 Rat Prote
53	245	9.1	1427	2	AAU10534	Aau10534 Human 160
54	244.5	9.0	909	3	AAB53448	Aab53448 Human col
55	244.5	9.0	1240	4	AAG67538	Aag67538 Amino aci
56	244	9.0	1202	3	AAB08521	Abb08521 A murine
57	244	9.0	1212	6	ABP97029	Abp97029 Rat L-FIL
58	243.5	9.0	693	6	ADA14410	Ada14410 Mouse spe
59	242	9.0	1484	2	AAW89721	Aaw89721 Canine ri
60	242	9.0	3259	7	ADE56037	Ade56037 Human pro
61	242	9.0	3259	7	ADE56033	Ade56033 Human pro
62	240	8.9	2633	4	ABG06505	Abg06505 Novel hum
63	239.5	8.9	1898	2	AAV30795	Aay30795 A human t
64	239.5	8.9	1898	7	ADD48869	Add48869 Human pro
65	239	8.8	2192	2	AAW21732	Aaw21732 LexA/NuMA
66	239	8.8	2272	2	AAW21731	Aaw21731 GAL4/HA/N
67	239	8.8	2663	4	AAW39097	Aam39097 Human pol
68	239	8.8	2688	4	AAW40883	Aam40883 Human pol
69	238	8.8	892	7	ADB79862	AdB79862 Rat myosi
70	238	8.8	892	7	ADD47859	Add47859 Rat prote
71	238	8.8	892	7	ADD47855	Add47855 Rat Prote
72	237.5	8.8	1213	4	AAW40016	Aam40016 Human pol
73	237.5	8.8	1213	6	ABP97031	Abp97031 Human L-F
74	236.5	8.8	1939	5	ABW77096	Abw77096 Human alp
75	236	8.7	1411	7	AAW02258	Aaw02258 Nucleolar
76	235.5	8.7	1939	7	ADD47857	Add47857 Human pro
77	235.5	8.7	1939	7	ADD47861	Add47861 Human pro
78	235.5	8.7	1939	7	ADD47033	Add47033 Human pro
79	234.5	8.7	650	7	ADB65482	AdB65482 Human pro
80	234.5	8.7	1235	5	ABW77435	Abw77435 Human tum
81	234.5	8.7	1235	7	ADD46509	Add46509 Human pro
82	234.5	8.7	1235	7	ADD48877	Add48877 Human pro
83	234.5	8.7	1646	4	ABB66368	Abb66368 Drosophil
84	234.5	8.7	1705	4	ABB71344	Abb71344 Drosophil
85	233	8.6	2115	3	AAV49937	Aay49937 Human NUM
86	233	8.6	2207	3	AAV432041	Aau32041 Novel hum
87	232	8.6	1935	7	ADD45207	Add45207 Rat Prote
88	231.5	8.6	1886	2	AAW54241	Aaw54241 Rattus no
89	231	8.5	1386	5	ABB57354	Abb57354 Mouse lsc
90	230.5	8.5	860	7	ADC38519	Adc38519 MHC A pro
91	230.5	8.5	3187	7	ADE56031	Ade56031 Rat Prote
92	230.5	8.5	3187	7	ADE56035	Ade56035 Rat Prote
93	230.5	8.5	3899	6	ABR92048	AbR92048 Human cer
94	230.5	8.5	3907	6	ABR92047	AbR92047 Human cer
95	230.5	8.5	3911	7	ADB67140	AdB67140 Human A-K
96	230.5	8.5	3917	6	ABR92050	AbR92050 Human cer
97	230.5	8.5	3925	6	ABR92049	AbR92049 Human cer
98	230	8.5	1948	4	ABG21233	Abg21233 Novel hum

99	229.5	8.5	907	4	ABU53073	Intracell
100	229.5	8.5	941	4	ABU53072	Intracell
101	229.5	8.5	953	4	ABU53069	Intracell
102	229.5	8.5	2228	7	ABR61599	Human gol
103	229.5	8.5	2230	6	ABU07445	Protein d
104	229.5	8.5	2230	7	ABR61600	Human gol
105	229.5	8.5	2250	7	ABR61601	Human gol
106	229.5	8.5	2252	7	ABR61602	Human gol
107	229	8.5	1453	4	AAM39213	Human pol
108	229	8.5	1469	4	AAM39214	Human pol
109	229	8.5	1935	5	ABG31649	Amino aci
110	229	8.5	1976	6	ABU07447	Protein d
111	229	8.5	1988	4	AAM41000	Human pol
112	229	8.5	1988	4	AAM40999	Human pol
113	229	8.5	2779	4	ABR62371	Drosophil
114	228.5	8.5	1781	6	ABR39830	Human SOV
115	228.5	8.5	1941	7	ADC26271	Human NCA
116	228	8.4	949	4	ABG15508	Novel hum
117	228	8.4	1372	2	AAM56473	Protein w
118	228	8.4	2383	5	ABG5631	Human bre
119	228	8.4	2442	3	AAY77575	Human cyc
120	227.5	8.4	660	4	ABR30817	Peptide #
121	227.5	8.4	660	5	ABG38772	Human pep
122	227.5	8.4	678	5	ABP41521	Human ova
123	227.5	8.4	1038	5	ABP73939	Candida a
124	227.5	8.4	1162	3	AAY96255	Kaposi's
125	227.5	8.4	1162	3	AAY58500	HRV8 ORF
126	227.5	8.4	1162	4	ABR62331	Amino aci
127	227.5	8.4	1162	5	ABR05621	Kaposi's
128	227.5	8.4	1521	3	AAG39233	Arabidops
129	227.5	8.4	1528	3	AAG39234	Arabidops
130	227.5	8.4	1562	3	AAG39233	Arabidops
131	227.5	8.4	1703	3	AAG36714	Arabidops
132	227.5	8.4	1710	3	AAG36713	Arabidops
133	227.5	8.4	1744	3	AAG36712	Arabidops
134	227	8.4	888	5	AAM30113	Human hom
135	226.5	8.4	688	6	ADA54360	Human pro
136	226.5	8.4	766	4	ABU53068	Transport
137	226.5	8.4	1790	6	ABR53116	Protein s
138	226.5	8.4	1940	5	ABG79661	Invertebr
139	226.5	8.4	3210	6	ABU07438	Protein d
140	226.5	8.4	3248	6	AAR9795	Kinetoch
141	225.5	8.3	1233	7	ADC33042	Human nov
142	225	8.3	923	4	ABG08600	Novel hum
143	225	8.3	1154	4	AAM32042	Novel hum
144	225	8.3	1857	3	AAY53970	Human per
145	224	8.3	1976	7	ADE63514	Rat Prote
146	224	8.3	1976	7	ADE63518	Rat Prote
147	224	8.3	2117	4	AAM32040	Novel hum
148	224	8.3	2246	4	ABG05850	Novel hum
149	223.5	8.3	484	4	AAM78985	Human pro
150	223.5	8.3	533	4	AAM79969	Human pro

ALIGNMENTS

RESULT 1	
AAW49042	
ID	AAW49042 standard; protein; 530 AA.
XX	
AC	AAW49042;
XX	
DT	09-NOV-1998 (first entry)
DE	Human low density lipoprotein binding protein LBP-3.
XX	
KW	Low density lipoprotein binding protein; LDL binding protein 3; LBP-3;
KW	receptor; human; atherosclerosis; diagnosis; therapy; vaccine.
XX	
OS	Homo sapiens.
XX	
FH	Key Location/Qualifiers

Peptide	53. 59	
	/note= "Claim 2"	
W09823282-A1.		
04-JUN-1998.		
26-NOV-1997;	97WO-US021857.	
27-NOV-1996;	96US-0031930P.	
03-JUN-1997;	97US-0048547P.	
(BOST-) BOSTON HEART FOUND INC.		
Lees AM, Lees RS, Law SW, Arjona AA;		
WPI: 1998-322455/28.		
N-PSDB; AAV32839.		
Nucleic acid encoding low density lipoprotein binding proteins and related vectors - transformed cells, proteins, and modulators of binding, useful for treatment and diagnosis of atherosclerosis and for identifying subjects at risk.		
Claim 1; Fig 8; 47pp; English.		
This polypeptide comprises novel human low density lipoprotein (LDL) binding protein LBP-3 that is capable of binding both native and methyl LDL. Its amino acid sequence was deduced from an isolated cDNA clone (see AAV32839). cDNA clones (see AAV32834-39) and encoded rabbit and human LBPs (see AAM49037-42) are claimed. An abnormality in an aspect of LBP metabolism or structure is diagnostic of a risk for atherosclerosis. The invention provides methods for determining if an animal is at risk for atherosclerosis (e.g. for prenatal screening); methods for treating atherosclerosis (including gene therapy) using e.g. LBP polypeptides to bind LDL and thereby prevent formation of atherosclerotic plaque; and methods for treating a cell having an abnormality in LBP structure or metabolism. Pharmaceutical and vaccine compositions are also provided, as well as recombinant vectors and host cells used to produce recombinant LBP		
Sequence 530 AA;		
Query Match	100.0%; Score 2702; DB 2; Length 530;	
Best Local Similarity	100.0%; Pred. No. 5e-160;	
Matches 530; Conservative 0; Mismatches 0; Indels 0; Gaps 0;		
QY	1 KSSPGQPEAGPEGAQERPSQAAPAVEAEGPGSSQAPRKEGAQAARTAQSGALRDVSEELS	60
Db	1 KSSPGQPEAGPEGAQERPSQAAPAVEAEGPGSSQAPRKEGAQAARTAQSGALRDVSEELS	60
QY	61 ROLDILSTYCYDNNQGGFEDGAQEPEDAEKSRITYVARNGEPEPTPVVYGEKPS	120
Db	61 ROLDILSTYCYDNNQGGFEDGAQEPEDAEKSRITYVARNGEPEPTPVVYGEKPS	120
QY	121 KGDPTNEIRQSDVEGDRDRRPPQKKKAGLKGKETTLMQTLNTLSTPEEKLAALCKKY	180
Db	121 KGDPTNEIRQSDVEGDRDRRPPQKKKAGLKGKETTLMQTLNTLSTPEEKLAALCKKY	180
QY	181 AELLLEHRNSQOMKLLQKQSQOLVQKDHLEHRSKAVLARSKLESICRELOHNRSLK	240
Db	181 AELLLEHRNSQOMKLLQKQSQOLVQKDHLEHRSKAVLARSKLESICRELOHNRSLK	240
QY	241 EGQVQAREEEERKEVTSHTSHFQVTLNDIQLQMEHNRNSKLRQENNELAERLKLIEQY	300
Db	241 EGQVQAREEEERKEVTSHTSHFQVTLNDIQLQMEHNRNSKLRQENNELAERLKLIEQY	300
QY	301 ELREEHIDKVFHKDLQQLQQLVDAKLQQAQEMLKAEERHOREKDFLLKEAVESQRMCELM	360
Db	301 ELREEHIDKVFHKDLQQLQQLVDAKLQQAQEMLKAEERHOREKDFLLKEAVESQRMCELM	360
QY	361 KQOETHLKOALALYTEKPEEFQNTLSKSEVFTTFKQEMHOMTKIKLLEKETTMYRSW	420

Db	361	KQETHLKQOLALYTEKFEFQNTLSKSSSEVFTTFKQEMEKMTKKIKKLEKETTMYSRW	420
Qy	421	ESSNKALLEMAEKTVDKLEGLQVKIORLEKLCRALOTERNDLNKRVDLSAGGQSSL	480
Db	421	ESSNKALLEMAEKTVDKLEGLQVKIORLEKLCRALOTERNDLNKRVDLSAGGQSSL	480
Qy	481	TDSGPERPFGGAQAPSSPRVTEAPCYFGAPSTEASGQTGPQETSARA	530
Db	481	TDSGPERPFGGAQAPSSPRVTEAPCYFGAPSTEASGQTGPQETSARA	530
RESULT 2			
ID	AAB82804	standard; protein; 530 AA.	
XX	XX		
DT	12-NOV-2001	(first entry)	
DE	XX		
XX	XX	Human low density lipoprotein binding protein 3 (LBP-3).	
KW	XX	Low density lipoprotein binding protein 3; LBP-3; LDL; human;	
KW	XX	atherosclerosis; antiarteriosclerotic; therapy; diagnosis; vaccine.	
OS	XX	Homo sapiens.	
PN	XX	WO200164874-A2.	
PD	XX	07-SEP-2001.	
XX	XX		
FF	XX	28-FEB-2001; 2001WO-US006356.	
PR	XX	02-MAR-2000; 2000US-00517849.	
PR	XX	14-JUL-2000; 2000US-00616289.	
XX	XX	(BOST-) BOSTON HEART FOUND INC.	
PI	XX	Lees AM, Lees RS, Law SW, Arjona AA;	
DR	XX	WPI; 2001-565505/63.	
DR	XX	N-PSDB; AAB26496.	
XX	XX		
PT	XX	New isolated low density lipoprotein binding polypeptide for treating,	
PT	XX	diagnosing and/or identifying therapeutic agents for atherosclerosis.	
PS	XX	Claim 13(h); Fig 8B; 143pp; English.	
XX	XX		
CC	XX	The present sequence is that of the N-terminal portion of novel human low	
CC	XX	density lipoprotein binding protein 3 (LBP-3). The amino acid sequence is	
CC	XX	deduced from an isolated partial cDNA clone (see AAB26494). A full-length	
CC	XX	sequence is given in AAB82808. Human LBP-3 is an example of claimed LBP	
CC	XX	polypeptides of the invention that are capable of binding to native and	
CC	XX	methylated low density lipoproteins. Also claimed are biologically active	
CC	XX	fragments and analogues of LBPs, polynucleotides encoding LBPs, as well	
CC	XX	as expression vectors, cells and methods of producing the LBPs. Methods	
CC	XX	of determining if an animal is at risk for atherosclerosis, methods for	
CC	XX	evaluating an agent for use in treating atherosclerosis, and methods for	
CC	XX	treating a cell having an abnormality in structure or metabolism of LBP	
CC	XX	are claimed. Pharmaceutical compositions comprising an LBP polypeptide or	
XX	XX	nucleic acid, and vaccine compositions, are also claimed	
SQ	XX	Sequence 530 AA;	
Query Match 100.0%; Score 2702; DB 4; Length 530;			
Best Local Similarity 100.0%; Pred. No. 5e-160;			
Matches 530; Conservative 0; Mismatches 0; Indels 0; Gaps 0;			
Qy	1	KSSPGQPEAGPEGAQRPSQAAPVAEAGPGSSQAQRKPEGAQARTAQSGALRDVSEILS	60
Db	1	KSSPGQPEAGPEGAQRPSQAAPVAEAGPGSSQAQRKPEGAQARTAQSGALRDVSEILS	60
Qy	61	ROLEEDILSTYCVNNQGGPDGAQGEPAEPDAEKSRITYVARNGEPPEPTFWYGEKPS	120

Db	61	ROLEEDILSTYCVNNQGGPDGAQGEPAEPDAEKSRITYVARNGEPPEPTFWYGEKPS	120
Qy	121	KGDPNTEIROSDEVDGDRHRPQPKKAKGLGKEITLLMOTLNTLSTPEKLAALCKKY	180
Db	121	KGDPNTEIROSDEVDGDRHRPQPKKAKGLGKEITLLMOTLNTLSTPEKLAALCKKY	180
Qy	181	AELLEHHNSOKMKLLQKQSLVQEKDHLRGEHSKAVLARSKLESLCRELQRNRSIK	240
Db	181	AELLEHHNSOKMKLLQKQSLVQEKDHLRGEHSKAVLARSKLESLCRELQRNRSIK	240
Qy	241	EEGVQARAEERKEKVTSHFPQVTLNDIQLQMEOHNRNSKLRQENMELARLKKLIQY	300
Db	241	EEGVQARAEERKEKVTSHFPQVTLNDIQLQMEOHNRNSKLRQENMELARLKKLIQY	300
Qy	301	ELREHIDKVPKHDLOQLVDKILQQAQEMLEKAEERHOREKDFLLKEAVESQRMCEIM	360
Db	301	ELREHIDKVPKHDLOQLVDKILQQAQEMLEKAEERHOREKDFLLKEAVESQRMCEIM	360
Qy	361	KQETHLKQOLALYTEKFEFQNTLSKSSSEVFTTFKQEMEKMTKKIKKLEKETTMYSRW	420
Db	361	KQETHLKQOLALYTEKFEFQNTLSKSSSEVFTTFKQEMEKMTKKIKKLEKETTMYSRW	420
Qy	421	ESSNKALLEMAEKTVDKLEGLQVKIORLEKLCRALOTERNDLNKRVDLSAGGQSSL	480
Db	421	ESSNKALLEMAEKTVDKLEGLQVKIORLEKLCRALOTERNDLNKRVDLSAGGQSSL	480
Qy	481	TDSGPERRPFGGAQAPSSPRVTEAPCYFGAPSTEASGQTGPQETSARA	530
Db	481	TDSGPERRPFGGAQAPSSPRVTEAPCYFGAPSTEASGQTGPQETSARA	530
RESULT 3			
ID	AAB82808	standard; protein; 546 AA.	
AC	AAB82808;		
XX	XX		
DT	12-NOV-2001	(first entry)	
XX	XX	Human low density lipoprotein binding protein 3 (LBP-3).	
KW	XX	Low density lipoprotein binding protein 3; LBP-3; LDL; human;	
KW	XX	atherosclerosis; antiarteriosclerotic; therapy; diagnosis; vaccine.	
OS	XX	Homo sapiens.	
PN	XX	WO200164874-A2.	
PD	XX	07-SEP-2001.	
XX	XX		
FF	XX	28-FEB-2001; 2001WO-US006356.	
PR	XX	02-MAR-2000; 2000US-00517849.	
PR	XX	14-JUL-2000; 2000US-00616289.	
XX	XX	(BOST-) BOSTON HEART FOUND INC.	
PI	XX	Lees AM, Lees RS, Law SW, Arjona AA;	
DR	XX	WPI; 2001-565505/63.	
DR	XX	N-PSDB; AAB26501.	
XX	XX		
PT	XX	New isolated low density lipoprotein binding polypeptide for treating,	
PT	XX	diagnosing and/or identifying therapeutic agents for atherosclerosis.	
PS	XX	Claim 13(l); Fig 8A; 143pp; English.	
XX	XX		
CC	XX	The present sequence is that of novel human low density lipoprotein	
CC	XX	binding protein 3 (LBP-3). The amino acid sequence was deduced from the	
CC	XX	coding region of isolated genomic DNA (see AAB26501). It differs from the	
CC	XX	sequence predicted from an LBP-3 cDNA clone (see AAB82804) by the	
CC	XX	presence of an additional 16 amino acids at the N-terminus (the cDNA	
CC	XX	clone is 5' truncated) and by having asparagine at amino acid position	

CC 130 rather than tyrosine. Human LBP-2 is an example of claimed LBP
CC polypeptides of the invention that are capable of binding to native and
CC methylated low density lipoproteins. Also claimed are biologically active
CC fragments and analogues of these LBPs, polynucleotides encoding LBPs, as
CC well as expression vectors, cells and methods of producing the LBPs. A
CC polypeptide having the amino acid residues 96-110 of the present sequence
CC is claimed (see A482820). Methods of determining if an animal is at risk
CC for atherosclerosis, methods for evaluating an agent for use in treating
CC atherosclerosis, and methods for treating a cell having an abnormality in
CC structure or metabolism of LBP are claimed. Pharmaceutical compositions
CC comprising an LBP polypeptide or nucleic acid, and vaccine compositions,
CC are also claimed
XX
SQ Sequence 546 AA;

Query Match 99.7%; Score 2693; DB 4; Length 546;
Best Local Similarity 99.8%; Pred. No. 1.9e-159;
Matches 529; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 KSPGQPEAGPEGAQRPSPQAAPVAEAGPGSSQAPRKPEGAQARTAQSGALDVSEELS 60
DB 17 KSPGQPEAGPEGAQRPSPQAAPVAEAGPGSSQAPRKPEGAQARTAQSGALDVSEELS 76

QY 61 RQLEDILSTVCVNNQGGPDEGAQGPAPPEPAEAKSRVTYVARGNPEPTPVVYGEKPS 120
DB 77 RQLEDILSTVCVNNQGGPDEGAQGPAPPEPAEAKSRVTYVARGNPEPTPVVYGEKPS 136

QY 121 KGPNTPEEIRQSDVGVDRDHRPQEKKAAGLKGITLLMOTLNTLSTPEEKLAALCKKY 180
DB 137 KGPNTPEEIRQSDVGVDRDHRPQEKKAAGLKGITLLMOTLNTLSTPEEKLAALCKKY 196

QY 181 AELLLEHNSOKMKLQKQSOVLQVQKHLRGEHSHKAVLARSLSLCLRELQHRNSLK 240
DB 197 AELLLEHNSOKMKLQKQSOVLQVQKHLRGEHSHKAVLARSLSLCLRELQHRNSLK 256

QY 241 EGVQARAREEERKEVTSHFQVTLNDIQLQMEQHNERNSKLRQENNELAERLKKLIBQY 300
DB 257 EGVQARAREEERKEVTSHFQVTLNDIQLQMEQHNERNSKLRQENNELAERLKKLIBQY 316

QY 301 ELBEHIDKVKFKHDKLOQQLVDKLOQAQEMKAEERHOREKDFLLKAVESQRCMLM 360
DB 317 ELBEHIDKVKFKHDKLOQQLVDKLOQAQEMKAEERHOREKDFLLKAVESQRCMLM 376

QY 361 KQDETHLKQALALYTFKFEFQNTLKSSEVFTTFQEMKMTKKIKLEKETTYRSRW 420
DB 377 KQDETHLKQALALYTFKFEFQNTLKSSEVFTTFQEMKMTKKIKLEKETTYRSRW 436

QY 421 ESSNKALLEVABEKTVPDKELGLQVKIQRLKLCRALQTERNDLNKRVQDLSAGQGSLL 480
DB 437 ESSNKALLEVABEKTVPDKELGLQVKIQRLKLCRALQTERNDLNKRVQDLSAGQGSLL 496

QY 481 TDSGPRRRPGGAQAPSPRVTEAPCYPGAPSTESGOTGPQEPPTSARA 530
DB 497 TDSGPRRRPGGAQAPSPRVTEAPCYPGAPSTESGOTGPQEPPTSARA 546

RESULT 4
ABR41210
ID ABR41210 standard; protein; 639 AA.
XX
AC ABR41210;
XX
DT 02-JUN-2003 (first entry)
XX
DE Human DITHP extracellular signalling protein.
XX
KW Human; dithp; diagnostic and therapeutic polynucleotide; diagnosis;
KW cancer; cell proliferative disorder; autoimmune disorder;
KW inflammatory disorder; infection; hormonal disorder; metabolic disorder;
KW neurological disorder; gastrointestinal disorder; transport disorder;
KW connective tissue disorder; drug screening; proteome analysis;
KW gene therapy; antisense therapy; genotyping; transgenic animal; knock in;
KW disease model; toxicological testing; transcript imaging;

XW extracellular signalling.
XX Homo sapiens.
XX WO200297031-A2.
XX 05-DEC-2002.
XX 27-MAR-2002; 2002WO-US010056.
XX 28-MAR-2001; 2001US-0279619P.
XX 29-MAR-2001; 2001US-0280067P.
XX 16-MAY-2001; 2001US-0280068P.
XX 17-MAY-2001; 2001US-0291280P.
XX 17-MAY-2001; 2001US-0291829P.
XX 19-JUN-2001; 2001US-0291849P.
XX 20-JUN-2001; 2001US-0299428P.
XX 20-JUN-2001; 2001US-0299776P.
XX 20-JUN-2001; 2001US-0300001P.
XX (INCY-) INCYTE GENOMICS INC.
XX Daffo A, Jones AL, Tran AB, Dahl CR, Gietzen D, Chinn J;
XX Dafour GE, Hillman JL, Yu JY, Tuason O, Yap PE, Amshey SR;
XX Doughtery SC, Dam TC, Liu TP, Nguyen DA, Kleefeld Y, Gerstin EH;
XX Peralta CH, David WH, Lewis SA, Chen AJ, Panzer SR, Harris B;
XX Flores V, Marwaha R, Lo A, Ian RY, Urashka ME;
XX WPI; 2003-129518/12.
XX N-PSDB; ACC46153.
XX Novel human diagnostic and therapeutic polypeptide useful for identifying
XX test compound which specifically binds to a polypeptide encoded by human
XX diagnostic and therapeutic polynucleotide, and to induce antibodies.
XX Claim 27; SEQ ID NO 745; 591pp; English.
XX The invention relates to novel human diagnostic and therapeutic
XX polynucleotides designated dithp (ACC46080-ACC46749) and to their encoded
XX proteins (DITHP: ABR41136-ABR41812). The invention also relates to
XX polynucleotide sequences at least 90% identical to the dithp cDNA
XX sequences of the invention; recombinant vectors, host cells and
XX transgenic organisms comprising a dithp nucleic acid sequence; the
XX recombinant production of DITHP proteins; antibodies specific for DITHP
XX proteins; microarrays comprising dithp nucleic acid sequences; methods of
XX detecting dithp nucleotide and protein sequences; methods of screening
XX for compounds which specifically bind a DITHP protein; and methods of
XX assessing the toxicity of test compounds using a dithp hybridisation
XX probe. Dithp nucleic acid sequences and DITHP proteins may be used in the
XX diagnosis of a wide variety of conditions including cancer and other cell
XX proliferative disorders; autoimmune or inflammatory disorders; bacterial,
XX viral, fungal or parasitic infections; hormonal disorders; metabolic
XX disorders; neurological disorders; gastrointestinal disorders; transport
XX disorders; and connective tissue disorders. They may also be used to
XX screen for modulators of protein activity or gene expression. DITHP
XX proteins can additionally be used in analysis of the proteome of a tissue
XX or cell type and to induce antibodies. The dithp nucleic acids are
XX additionally useful in somatic or germline gene therapy of the disorders
XX mentioned above, as a source of antisense sequences, as a source of
XX probes and primers, in genotyping and identification of individuals, in
XX the generation of transgenic animal models of human disease or knock in
XX humanised animals, in toxicological testing, and in transcript imaging.
XX The present sequence represents a DITHP protein which has extracellular
XX signalling activity. Note: The sequence data for this patent did not form
XX part of the printed specification, but was obtained in electronic format
XX directly from WIPO at ftp.wipo.int/pub/published_pct_sequences
XX Sequence 639 AA;

Query Match 99.7%; Score 2693; DB 6; Length 639;
Best Local Similarity 99.8%; Pred. No. 2.2e-159;
Matches 529; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 KSSPQPEAGPEGAQERPSQAAPAVEAGPGSSQAPRKPEGAQARTAQSGALRDVSEELS 60
DB 110 KSSPQPEAGPEGAQERPSQAAPAVEAGPGSSQAPRKPEGAQARTAQSGALRDVSEELS 169
QY 61 ROLEDILSTYCVNNQGGEDGAGGAEPAEDAEKSTYVARNGEPEPTPVVYGEKPS 120
DB 170 ROLEDILSTYCVNNQGGEDGAGGAEPAEDAEKSTYVARNGEPEPTPVVYGEKPS 229
QY 121 KGDPTNTEIROSDVGDHRRPQPKKAGKGLGKEITLLMOTLNTLSTPEEKLAALCKKY 180
DB 230 KGDPTNTEIROSDVGDHRRPQPKKAGKGLGKEITLLMOTLNTLSTPEEKLAALCKKY 289
QY 181 AELLEHRNSQOMKLLQKQSQVLQVQKHRLRGHSHKAVLARSKLESICRELQHRNLSLK 240
DB 290 AELLEHRNSQOMKLLQKQSQVLQVQKHRLRGHSHKAVLARSKLESICRELQHRNLSLK 349
QY 241 EGVQARAREEERKEVTSHFQVTLNDIQLQMEQHNRNRSKLQENMELARLKLIEQY 300
DB 350 EGVQARAREEERKEVTSHFQVTLNDIQLQMEQHNRNRSKLQENMELARLKLIEQY 409
QY 301 ELREBHIDKVPKHKDLQOOLVDKLAQQAQEMLKEAEERHOREKDFLLKEAVESQRMCELM 360
DB 410 ELREBHIDKVPKHKDLQOOLVDKLAQQAQEMLKEAEERHOREKDFLLKEAVESQRMCELM 469
QY 361 KQOETHLQOALYKTEKPEEFONTLSKSSEVFTTFQEMEKMTKIKKLEKETTMYRSRW 420
DB 470 KQOETHLQOALYKTEKPEEFONTLSKSSEVFTTFQEMEKMTKIKKLEKETTMYRSRW 529
QY 421 ESNVALLMABEKTVDKLEGLQVKIQRLEKLCRALQTERNDLNKRVQDLSAGQGSGL 480
DB 530 ESNVALLMABEKTVDKLEGLQVKIQRLEKLCRALQTERNDLNKRVQDLSAGQGSGL 589
QY 481 TDSGPERPEGCAQAPSPRVTEAPCPYGPAPSTASGOTGPCEPTSARA 530
DB 590 TDSGPERPEGCAQAPSPRVTEAPCPYGPAPSTASGOTGPCEPTSARA 639

RESULT 5
ID ABR41186
AC ABR41186 standard; protein; 639 AA.
XX ABR41186;
XX 02-JUN-2003 (first entry)
XX Human DITHP extracellular signalling protein.
XX Human; dithp; diagnostic and therapeutic polynucleotide; diagnosis;
XX cancer; cell proliferative disorder; autoimmune disorder;
XX inflammatory disorder; infection; hormonal disorder; metabolic disorder;
XX neurological disorder; gastrointestinal disorder; transport disorder;
XX connective tissue disorder; drug screening; proteome analysis;
XX gene therapy; antisense therapy; genotyping; transgenic animal; knock in;
XX disease model; toxicological testing; transcript imaging;
XX extracellular signalling.
XX Homo sapiens.
XX WO200297031-A2.
XX 05-DEC-2002.
XX 27-MAR-2002; 2002WO-US010056.
XX 28-MAR-2001; 2001US-0279619P.
XX 29-MAR-2001; 2001US-0280067P.
XX 29-MAR-2001; 2001US-0280068P.
XX 16-MAY-2001; 2001US-0291280P.
XX 17-MAY-2001; 2001US-0291829P.
XX 17-MAY-2001; 2001US-0291849P.
XX 19-JUN-2001; 2001US-0299428P.
XX 20-JUN-2001; 2001US-0299776P.
XX 20-JUN-2001; 2001US-0300001P.

XX (INCY-) INCYTE GENOMICS INC.
XX Daffo A, Jones AL, Tran AB, Dahl CR, Gietzen D, Chinn J;
XX Dufour GE, Hillman JL, Yu JY, Tuason O, Yap PE, Amshey SR;
XX Daugherty SC, Dam TC, Liu TF, Nguyen DA, Klesfeld Y, Gerstein EH;
XX Peralta CH, David MH, Lewis SA, Chen AJ, Panzer SR, Harris B,
XX Flores V, Marwaha R, Lo A, Lan RY, Urashka ME;
XX WPI; 2003-129518/12.
XX N-PSDB; ACC46130.
XX Novel human diagnostic and therapeutic polypeptide useful for identifying
XX test compound which specifically binds to a polypeptide encoded by human
XX diagnostic and therapeutic polynucleotide, and to induce antibodies.
XX Claim 27; SEQ ID NO 721; 591bp; English.
XX The invention relates to novel human diagnostic and therapeutic
XX polynucleotides designated dithp (ACC46080-ACC46749) and to their encoded
XX proteins (DITHP; ABR41136-ABR41812). The invention also relates to
XX polynucleotide sequences at least 90% identical to the dithp cDNA
XX sequences of the invention; recombinant vectors, host cells and
XX transgenic organisms comprising a dithp nucleic acid sequence; the
XX recombinant production of DITHP proteins; antibodies specific for DITHP
XX proteins; microarrays comprising dithp nucleic acid sequences; methods of
XX detecting dithp nucleotide and protein sequences; methods of screening
XX for compounds which specifically bind a DITHP protein; and methods of
XX assessing the toxicity of test compounds using a dithp hybridisation
XX probe. Dithp nucleic acid sequences and DITHP proteins may be used in the
XX diagnosis of a wide variety of conditions including cancer and other cell
XX proliferative disorders; autoimmune or inflammatory disorders; bacterial,
XX viral, fungal or parasitic infections; hormonal disorders; metabolic
XX disorders; neurological disorders; gastrointestinal disorders; transport
XX disorders; and connective tissue disorders. They may also be used to
XX screen for modulators of protein activity or gene expression. DITHP
XX proteins can additionally be used in analysis of the proteome of a tissue
XX or cell type and to induce antibodies. The dithp nucleic acids are
XX additionally useful in somatic or germline gene therapy of the disorders
XX mentioned above, as a source of antisense sequences, as a source of
XX probes and primers, in genotyping and identification of individuals, in
XX the generation of transgenic animal models of human disease or knock in
XX humanised animals, in toxicological testing, and in transcript imaging.
XX The present sequence represents a DITHP protein which has extracellular
XX signalling activity. Note: The sequence data for this patent did not form
XX part of the printed specification, but was obtained in electronic format
XX directly from WIPO at ftp.wipo.int/pub/published_pct_sequences
XX Sequence 639 AA;
XX Query Match 99.7%; Score 2693; DB 6; Length 639;
XX Best Local Similarity 99.8%; Pred. No. 2.2e-159;
XX Matches 529; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 1 KSSPQPEAGPEGAQERPSQAAPAVEAGPGSSQAPRKPEGAQARTAQSGALRDVSEELS 60
DB 110 KSSPQPEAGPEGAQERPSQAAPAVEAGPGSSQAPRKPEGAQARTAQSGALRDVSEELS 169
QY 61 ROLEDILSTYCVNNQGGEDGAGGAEPAEDAEKSTYVARNGEPEPTPVVYGEKPS 120
DB 170 ROLEDILSTYCVNNQGGEDGAGGAEPAEDAEKSTYVARNGEPEPTPVVYGEKPS 229
QY 121 KGDPTNTEIROSDVGDHRRPQPKKAGKGLGKEITLLMOTLNTLSTPEEKLAALCKKY 180
DB 230 KGDPTNTEIROSDVGDHRRPQPKKAGKGLGKEITLLMOTLNTLSTPEEKLAALCKKY 289
QY 181 AELLEHRNSQOMKLLQKQSQVLQVQKHRLRGHSHKAVLARSKLESICRELQHRNLSLK 240
DB 290 AELLEHRNSQOMKLLQKQSQVLQVQKHRLRGHSHKAVLARSKLESICRELQHRNLSLK 349
QY 241 EGVQARAREEERKEVTSHFQVTLNDIQLQMEQHNRNRSKLQENMELARLKLIEQY 300
DB 350 EGVQARAREEERKEVTSHFQVTLNDIQLQMEQHNRNRSKLQENMELARLKLIEQY 409

QY	301	ELREEHIDKVFHKDLQOQLVDKLAQQAQEMKAEABERHOREKDFLLKEAVESORMCELM	360
DB	410	ELREEHIDKVFHKDLQOQLVDKLAQQAQEMKAEABERHOREKDFLLKEAVESORMCELM	469
QY	361	KQOETHLKQOLALYTKRPFQNTLSKSEVFTTFQEWKMTKKIKLEKETTMYRSRW	420
DB	470	KQOETHLKQOLALYTKRPFQNTLSKSEVFTTFQEWKMTKKIKLEKETTMYRSRW	529
QY	421	ESSNKALLEMAEKTVDKLEGLQVKIQRLEKLCRALQTERNDLNKRVQDLSAGGQSSL	480
DB	530	ESSNKALLEMAEKTVDKLEGLQVKIQRLEKLCRALQTERNDLNKRVQDLSAGGQSSL	589
QY	481	TDGSPRRREGGAQAPSSPRVTEAPCYGAPSTASGOTGQPEPTSA	530
DB	590	TDGSPRRREGGAQAPSSPRVTEAPCYGAPSTASGOTGQPEPTSA	639
RESULT 6			
ID	AAW49039	standard; protein; 557 AA.	
XX	AAW49039;		
DT	09-NOV-1998	(first entry)	
DE	Rabbit low density lipoprotein binding protein LBP-3.		
KW	Low density lipoprotein binding protein; LDL binding protein 3; LBP-3;		
KW	receptor; rabbit; atherosclerosis; diagnosis; therapy; vaccine.		
OS	Oryctolagus cuniculus.		
Key	Location/Qualifiers		
Peptide	96..110 /note="Claim 2"		
WO9823282-A1.			
04-JUN-1998.			
26-NOV-1997;	97WO-US021857.		
27-NOV-1996;	96US-0031930P.		
03-JUN-1997;	97US-0048547P.		
(BOST-) BOSTON HEART FOUND INC.			
Lees AM, Lees RS, Law SM, Arjona AA;			
WPI; 1998-322455/28.			
N-PSDB; AAV32836.			
Nucleic acid encoding low density lipoprotein binding proteins and			
related vectors - transformed cells, proteins, and modulators of binding,			
useful for treatment and diagnosis of atherosclerosis and for identifying			
subjects at risk.			
Claim 1; Fig 5; 47pp; English.			
This polypeptide comprises novel rabbit low density lipoprotein (LDL)			
binding protein LBP-3 that is capable of binding both native and methyl			
LDL. Its amino acid sequence was deduced from rabbit abdominal aorta cDNA			
(see AAV32836). cDNA clones (see AAV32834-39) and encoded rabbit and human			
LBPs (see AAW49037-42) are claimed. An abnormality in an aspect of LBP			
metabolism or structure is diagnostic of a risk for atherosclerosis. The			
invention provides: methods for determining if an animal is at risk for			
atherosclerosis (e.g. for prenatal screening); methods for treating			
atherosclerosis (including gene therapy) using e.g. LBP polypeptides to			
bind LDL and thereby prevent formation of atherosclerotic plaque; and			
methods for treating a cell having an abnormality in LBP structure or			
metabolism. Pharmaceutical and vaccine compositions are also provided, as			
well as recombinant vectors and host cells used to produce recombinant			

CC	LBP																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																												
----	-----	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--

CC	LBP		
XX	Sequence 557 AA;		
QY	Query Match	91.2%; Score 2465.5; DB 2; Length 557;	
	Best Local Similarity	90.8%; Pred. No. 2.8e-145;	
	Matches 492; Conservative	12; Mismatches 25; Indels 13; Gaps 3;	
QY	1	KSSPCQEPGEGQAQERPSQAAPAVAEAGPGSSQAAPRKEGAQAQTAQAGALRDVSEELS	60
DB	17	KSSPCQEPGAGAGAGRPRPAPAREAG-ASSQAQRPGEQAQTAQAGALCDVSEELS	75
QY	61	QLEEDILSTYCVDNNGQGGPDGAGQEPAPEDAEKSRITYVARNGSEPP-TPVVYGEKEP	119
DB	76	QLEEDILSTYCVDNNGQAGPDGQVQEPPEDEAKSRAYVARNGEPGTPVWNGEKEP	135
QY	120	SKGPNTEEIROSDVEGDRDHRPQEKKAKGLKEITLLMCTLNTLSTPEKLAALCKK	179
DB	136	SKAPPGTSEIRTSDEWGRDHRFPQEKKAKGLKEITLLMCTLNTLSTPEKLAALCKK	195
QY	180	YAEILLEHNSQKQMKLLQKQSQQLVQEKDHLRGEHSAVLARSKLESCLRELQHNRS	239
DB	196	YAEILLEHNSQKQMKLLQKQSQQLVQEKDHLRGEHSAVLARSKLESCLRELQHNRS	255
QY	240	KEGVQARAREEBEKKREVTSHFOVTIANDIQLOMEQHNRNSKLRQENMELAEKLLIQ	299
DB	256	KEGVQARAREEBEKKREVTSHFOVTIANDIQLOMEQHNRNSKLRQENMELAEKLLIQ	315
QY	300	YELREHIDKVFHKDLQOQLVDKLAQQAQEMKAEABERHOREKDFLLKEAVESORMCEL	359
DB	316	YELREHIDKVFHKDLQOQLVDKLAQQAQEMKAEABERHOREKDFLLKEAVESORMCEL	375
QY	360	MKQETHLKQOLALYTKRPFQNTLSKSEVFTTFQEWKMTKKIKLEKETTMYRSR	419
DB	376	MKQETHLKQOLALYTKRPFQNTLSKSEVFTTFQEWKMTKKIKLEKETTMYRSR	435
QY	420	WESSNKALLEMAEKTVDKLEGLQVKIQRLKLCRALQTERNDLNKRVQDLSAGGQGS	479
DB	436	WESSNKALLEMAEKTVDKLEGLQVKIQRLKLCRALQTERNDLNKRVQDLSAGGQGP	495
QY	480	LTDSGPERRP-----EGPGAQAPSSPRVTEAPCYGAPSTASGOTGQPEPTSA	528
DB	496	VSDSGPERRPAPATTSKEQVGGGAQVPSNPATDASCAGAPSTASGOTGQPEPTTA	555
QY	529	RA 530	
DB	556	TA 557	
RESULT 7			
ID	AAB82801	standard; protein; 557 AA.	
XX	AAB82801;		
AC	AAB82801;		
DT	12-NOV-2001	(first entry)	
XX	Rabbit low density lipoprotein binding protein 3 (LBP-3).		
DE	Low density lipoprotein binding protein 3; LBP-3; LDL; rabbit;		
KW	atherosclerosis; antiarteriosclerotic; therapy; diagnosis; vaccine.		
OS	Oryctolagus cuniculus.		
PN	WO200164874-A2.		
XX	07-SEP-2001.		
PD	28-FEB-2001; 2001WO-US0063356.		
XX	02-MAR-2000; 2000US-00517849.		
PR	14-JUL-2000; 2000US-00616289.		
XX			

PA (BOST-) BOSTON HEART FOUND INC.
XX
PI Lees AM, Lees RS, Law SW, Arizona AA;
XX
DR WPI: 2001-565505/63.
DR N-PSDB; AAH26491.
XX
PT New isolated low density lipoprotein binding polypeptide for treating,
PT diagnosing and/or identifying therapeutic agents for atherosclerosis.
XX
PS Claim 13(e); Fig 5; 143pp; English.
XX
XX The present sequence is that of novel rabbit low density lipoprotein
CC binding protein 1 (LBP-3). The amino acid sequence is deduced from an
CC isolated cDNA clone (see AAH26491). Rabbit LBP-3 is an example of claimed
CC polypeptides of the invention, termed LBPs, that are capable of binding
CC to native and methylated low density lipoproteins. Also claimed are
CC biologically active fragments and analogues of LBPs, polynucleotides
CC encoding LBPs, as well as expression vectors, cells and methods of
CC producing the LBPs. A polypeptide having the amino acid residues 96-110
CC of the present sequence is claimed (see AAH26491). Methods of determining
CC if an animal is at risk for atherosclerosis, methods for evaluating an
CC agent for use in treating atherosclerosis, and methods for treating a
CC cell having an abnormality in structure or metabolism of LBP are claimed.
CC Pharmaceutical compositions comprising an LBP polypeptide or nucleic
CC acid, and vaccine compositions, are also claimed
XX
XX Sequence 557 AA;
XX
XX Query Match 91.2%; Score 2465.5; DB 4; Length 557;
XX Best Local Similarity 90.8%; Pred. No. 2.8e-145;
XX Matches 492; Conservative 12; Mismatches 25; Indels 13; Gaps 3;
XX
QY 1 KSSPGQPEAGPAGQERPSQAAPVAVAEAGPGSSQAPRKPEGAQTAQSGALRDVSELS 60
Db 17 KSSPGQPEAGPAGQERPSQAAPVAVAEAGPGSSQAPRKPEGAQTAQSGALRDVSELS 75
XX
QY 61 RQLEDILSTYCVNNNGGPGEDGAQGEPAEPEDAESRTYVARNGEPEP-TPVYGEKEP 119
Db 76 RQLEDILSTYCVNNNGGPGEDGAQGEPAEPEDAESRTYVARNGEPEP-TPVYGEKEP 135
XX
QY 120 SKGDPNTEIRSDVGRDHRPQKKKAGLGEITLLMTQTLNTLSTPEKLAALCKK 179
Db 136 SKAEPGTEIRTSDEVGDRDHRPQKKKAGLGEITLLMTQTLNTLSTPEKLAALCKK 195
XX
QY 180 YAEELLEHNRSOKMQLQKQSQVQLVQERDHLRGEHSKAVLARSKLESICRELQRHNSL 239
Db 196 YAEELLEHNRSOKMQLQKQSQVQLVQERDHLRGEHSKAVLARSKLESICRELQRHNSL 255
XX
QY 240 KEEGVORABEEKREKVTSHFQVTLNDIQLQWEQHNSKLRQENMELAEKLIKLEQ 299
Db 256 KEEGVORABEEKREKVTSHFQVTLNDIQLQWEQHNSKLRQENMELAEKLIKLEQ 315
XX
QY 300 YELREEHIDKVPFKHDLQQLVDALQQAQEMLKEAEERHQRKDFLLKEAVESQRMCEL 359
Db 316 YELREEHIDKVPFKHDLQQLVDALQQAQEMLKEAEERHQRKDFLLKEAVESQRMCEL 375
XX
QY 360 MKQOETHLQQAALYTERKEFEFNTLSKSEVFTTPKQENKQTKIKLEKETTMYRSR 419
Db 376 MKQOETHLQQAALYTERKEFEFNTLSKSEVFTTPKQENKQTKIKLEKETTMYRSR 435
XX
QY 420 WESSKALLEMAEKTVRDKLEGLQVKLRLEKCRALQTERNDLNKRVQDLSSAGQGS 479
Db 436 WESSKALLEMAEKTVRDKLEGLQVKLRLEKCRALQTERNDLNKRVQDLSSAGQGS 495
XX
QY 480 LTDSGPERP-----EGPGAQAPSSPRVTEAPCYVPGAPSTEASGQTGPQEPSTA 528
Db 496 VSDSGPERPAPATTSKQGVGPGAGVFNPRATDASCAGAPSTEASGQTGPQEPSTA 555
XX
QY 529 RA 530
Db 556 TA 557

RESULT 8

ABG00839
ID ABG00839 standard; protein; 1749 AA.

XX AC ABG00839;

XX DT 13-FEB-2002 (first entry)

XX DE Novel human diagnostic protein #830.

XX KW Human; chromosome mapping; gene mapping; gene therapy; forensic;
XX food supplement; medical imaging; diagnostic; genetic disorder.

XX OS Homo sapiens.

XX FN WO200175067-A2.

XX PD 11-OCT-2001.

XX PF 30-MAR-2001; 2001WO-US008631.

XX PR 31-MAR-2000; 2000US-00540217.

XX PR 23-AUG-2000; 2000US-00649167.

XX PA (HYSB-) HYSBQ INC.

XX PI Dmanac RT, Liu C, Tang YT;

XX DR WPI: 2001-639362/73.

XX DR N-PSDB; AAS65026.

XX PT New isolated polynucleotide and encoded polypeptides, useful in
XX diagnostics, forensics, gene mapping, identification of mutations
XX responsible for genetic disorders or other traits and to assess
XX biodiversity.

XX PS Claim 20; SEQ ID NO 31198; 103pp; English.

XX The invention relates to isolated polynucleotide (I) and polypeptide (II)
XX sequences. (I) is useful as hybridisation probes, polymerase chain
XX reaction (PCR) primers, oligomers, and for chromosome and gene mapping,
XX and in recombinant production of (II). The polynucleotides are also used
XX in diagnostics as expressed sequence tags for identifying expressed
XX genes. (I) is useful in gene therapy techniques to restore normal
XX activity of (II) or to treat disease states involving (II). (II) is
XX useful for generating antibodies against it, detecting or quantitating a
XX polypeptide in tissue, as molecular weight markers and as a food
XX supplement. (II) and its binding partners are useful in medical imaging
XX of sites expressing (II). (I) and (II) are useful for treating disorders
XX involving aberrant protein expression or biological activity. The
XX polypeptide and polynucleotide sequences have applications in
XX diagnostics, forensics, gene mapping, identification of mutations
XX and to produce other types of data and products dependent on DNA and
XX amino acid sequences. ABG0010-ABG30377 represent novel human diagnostic
XX amino acid sequences of the invention. Note: The sequence data for this
XX patent did not appear in the printed specification, but was obtained in
XX electronic format directly from WIPO at
XX ftp.wipo.int/pub/published_pct_sequences

XX SQ Sequence 1749 AA;

Query Match 86.5%; Score 2338.5; DB 4; Length 1749;

Best Local Similarity 84.3%; Pred. No. 8.4e-137;

Matches 483; Conservative 4; Mismatches 31; Indels 55; Gaps 6;

QY 1 KSSPGQPEAGPAGQERPSQAAPVAVAEAGPGSSQAPRKPEGAQTAQSGALRDVSELS 60

Db 168 KSSPGQPEAGPAGQERPSQAAPVAVAEAGPGSSQAPRKPEGAQTAQSGALRDVSELS 227

QY 61 RQLEDILSTYCVNNNGGPGEDGAQGEPAEPEDAESRTYVARNGEPEP-TPVYGEKEPS 120

Db 228 RQLEDILSTYCVNNOGGPGEAQPAPEDAEKSRVTVAENGPEPTPVVNGEKPS 287
QY 121 KGDPTTEIRQSDVEGDRHRRPQEKKAAGL-----GKETTLMQTLNTLSTPE 170
Db 288 KGDPTTEIRQSDVEGDRHRRPQEKKAAGLGEQRAALCEAGKEITLLMQTLNTLSTPE 347
QY 171 EKLAALCKKVAHLEHRNSQKMLLOKQSOVLQKDHLEHSHKAVLARSKLSLCKR 230
Db 348 EKLAALCKKVAHLEHRNSQKMLLOKQSOVLQKDHLEHSHKAVLARSKLSLCKR 407
QY 231 ELQRHNRSLK-----HEGVQARBEERKEKVTSHFQVTLNDIOL 270
Db 408 ELQRHNRSLKGDPCVAVDDGSRADPFAQEGVQARBEERKEKVTSHFQVTLNDIOL 467
QY 271 QMEQHRNSKLOENMELARLKLLEQVELBEHIDKFKHKLQOOLVDAKLOQAOE 330
Db 468 QMEQHRNSKLOENMELARLKLLEQVELBEHIDKFKHKLQOOLVDAKLOQAOE 527
QY 331 MKAEERHOREKDFLLKEAVESQRMCELMKQOETHLQOALYTEKFEFQNTLSKSE 390
Db 528 MKAEERHOREKDFLLKEAVESQRMCELMKQOETHLQOALYTEKFEFQNTLSKSE 587
QY 391 VFTTFQKEMKTKIKKLEKETTMYRSRWSKALLEMAEKTVDKLEGLQVKIOR 450
Db 588 VFTTFQKEMKTKIKKLEKETTMYRSRWSKALLEMAEKTVDKLEGLQVKIOR 647
QY 451 LEKLCRALQTERNDLNRVQDL-----SAGGCGSLTDGSPRRPEGCAQAP 497
Db 648 LEKLCRALQTERNDLNRVQDL-----SAGGCGSLTDGSPRRPEGCAQAP 707
QY 498 SS-----PRVTEAPCYPGAPSTASQGTGPQP 525
Db 708 TSARKRPNR---KPLLPG---DGEGLRPSQRP 733

RESULT 9
ABBI1764
ID ABBI1764 standard; peptide; 510 AA.
XX AC ABBI1764;
XX DT 11-JAN-2002 (first entry)
XX DE Human LDL binding protein homologue, SEQ ID NO:2134.
XX KW Human; cytokine; cell proliferation; cell differentiation; growth factor;
KW haematopoiesis regulation; tissue growth; immunomodulator; activin;
KW inhibin; chemotaxis; chemokinesis; thrombolysis; oncogenesis;
KW proliferation; metastasis; cancer; tumour; haematopoietic disorder;
KW myeloid cell disorder; lymphoid cell disorder; asthma; arthritis;
KW chronic inflammatory condition; proliferative retinopathy;
KW atherosclerosis; coronary heart disease; arterial ischaemia;
KW bone disorder; osteoporosis; vascular growth disorder;
KW tissue regeneration; wound healing; infection; immune disorder;
KW cell culture; drug screening; gene therapy; antiinflammatory;
KW antiasthmatic; antiarthritis; haemostatic; antiarteriosclerotic;
KW cytosstatic; osteopathic; vasotropic; cardiant; virucide; antibacterial;
KW antifungal; vulnerary; antiulcer.
XX OS Homo sapiens.
XX PN WO200157188-A2.
XX PD 09-AUG-2001.
XX PF 05-FEB-2001; 2001WO-00039800.
XX PR 03-FEB-2000; 2000US-00496914.
XX PR 27-APR-2000; 2000US-00560875.
XX XX (HYSE-) HYSEQ INC.
XX PA Tang YT, Liu C, Drmanac RT;
XX PI

XX MPI, 2001-457740/49.
DR N-PSDB; ABA09008.
XX Human proteins and DNA encoding sequences useful for preventing, treating
PT or ameliorating a medical condition in a mammalian subject e.g. arthritis
PT and cancer.
XX Claim 20; Page 240-241; 1963pp; English.
XX Sequences ABB10991-ABB21330 represent 1350 novel human polypeptides, and
CC sequences ABA08225-ABA09574 represent nucleic acids encoding them. The
CC invention also relates to vectors and recombinant host cells comprising a
CC nucleotide of the invention, methods of producing the novel polypeptides,
CC antibodies against the polypeptides, methods of detecting the nucleotides
CC or polypeptides in a sample, and methods of identifying compounds which
CC bind to polypeptides of the invention. Although novel, many of the
CC polypeptides of the invention have homology to known proteins, thereby
CC giving an insight into their probable biological activities, and hence
CC potential therapeutic applications. The polypeptides of the invention may
CC have various activities, including cytokine, cell proliferation or cell
CC differentiation activities; stem cell growth factor activity;
CC haematopoiesis regulatory activity; tissue growth activity;
CC immunomodulatory activity; activin- or inhibin-related activities;
CC chemotactic or chemokinetic activities; haemostatic, thrombotic or
CC thrombolytic activities; receptor or ligand activities; or may be
CC involved in oncogenesis, cancer cell proliferation or metastasis.
CC Depending on their biological activities, polypeptides and nucleotides of
CC the invention are useful for preventing, treating or ameliorating medical
CC conditions, e.g. by protein or gene therapy. Such conditions include
CC cancers, haematopoietic disorders (e.g. myeloid or lymphoid cell
CC disorders), chronic inflammatory conditions (e.g. asthma or arthritis),
CC proliferative retinopathy, atherosclerosis, coronary heart disease,
CC arterial ischaemia, bone disorders (e.g., osteoporosis), and abnormal
CC vascular growth. Polypeptides involved with tissue regeneration and
CC repair (or nucleic acids encoding them) may be used to promote wound
CC healing (e.g., of burns, incisions and ulcers), while those with
CC immunomodulatory activities may be used in the treatment of viral,
CC bacterial and fungal infections in addition to immune disorders.
CC Polypeptides with growth factor activity may be used in cell cultures to
CC promote cell growth. For example, such polypeptides may be used to
CC manipulate stem cells in culture to give rise to neuroepithelial cells
CC that can be used to augment or replace cells damaged by illness,
CC autoimmune disease or accidental damage. The polypeptides and nucleotides
CC may also be used in the diagnosis of the above conditions, and in drug
CC screening techniques. The present sequence represents a novel human
CC polypeptide of the invention
XX SQ Sequence 510 AA;

Query Match 86.0%; Score 2324; DB 4; Length 510;
Best Local Similarity 99.8%; Pred. No. 1.6e-136;
Matches 459; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 1 KSSPGQPEAGPEGAQERPSQAAPAVEAGFGSSQAPRKPEGAQARTAQSGALDVSEELS 60
Db 28 KSSPGQPEAGPEGAQERPSQAAPAVEAGFGSSQAPRKPEGAQARTAQSGALDVSEELS 87
QY 61 RQLEDILSTYCVNNOGGPGEAQPAPEDAEKSRVTVAENGPEPTPVVNGEKPS 120
Db 88 RQLEDILSTYCVNNOGGPGEAQPAPEDAEKSRVTVAENGPEPTPVVNGEKPS 147
QY 121 KGDPTTEIRQSDVEGDRHRRPQEKKAAGLKEITLLMQTLNTLSTPEKLAALCKKY 180
Db 148 KGDPTTEIRQSDVEGDRHRRPQEKKAAGLKEITLLMQTLNTLSTPEKLAALCKKY 207
QY 181 AEILLEHRNSQKMLLOKQSOVLQKDHLEHSHKAVLARSKLSLCKRLOHNRSLK 240
Db 208 AEILLEHRNSQKMLLOKQSOVLQKDHLEHSHKAVLARSKLSLCKRLOHNRSLK 267
QY 241 EGVQARBEERKEKVTSHFQVTLNDIOLQMEQHRNSKLOENMELARLKLLEQVEL 300
Db 268 EGVQARBEERKEKVTSHFQVTLNDIOLQMEQHRNSKLOENMELARLKLLEQVEL 327

QY 301 ELREHIDKVFHKDLQQLVDKLAQQAQEMKAEABERHOREKDFLLKEAVESQRMCELM 360
 DB 328 ELREHIDKVFHKDLQQLVDKLAQQAQEMKAEABERHOREKDFLLKEAVESQRMCELM 387
 QY 361 KQETHLKQQLALYTEKPEFQNTLSKSSSEVFTTFKQEMERQTKIKKLEKETTMYSRW 420
 DB 388 KQETHLKQQLALYTEKPEFQNTLSKSSSEVFTTFKQEMERQTKIKKLEKETTMYSRW 447
 QY 421 ESSNKALLEMAEBKTVRDKEGLQVKIQRLEKLCRALQT 460
 DB 448 ESSNKALLEMAEBKTVRDKEGLQVKIQRLEKLCRALQT 487

RESULT 10
 ID AAM79741
 ID AAM79741 standard; protein; 510 AA.
 AC AAM79741;
 DT 06-NOV-2001 (first entry)
 DE Human protein SEQ ID NO 3387.
 KW Human; cytokine; cell proliferation; cell differentiation; gene therapy;
 KW vaccine; peptide therapy; stem cell growth factor; haematopoiesis;
 KW tissue growth factor; immunomodulatory; cancer; leukaemia;
 KW nervous system disorder; arthritis; inflammation.
 OS Homo sapiens.
 QY WO200157190-A2.
 FN 09-AUG-2001.
 XX 05-FEB-2001; 2001WO-US004098.
 XX 03-FEB-2000; 2000US-00496914.
 XX 27-APR-2000; 2000US-00560875.
 XX 20-JUN-2000; 2000US-00598075.
 XX 19-JUL-2000; 2000US-00620325.
 XX 01-SEP-2000; 2000US-00654936.
 XX 15-SEP-2000; 2000US-00663561.
 XX 20-OCT-2000; 2000US-00693325.
 XX 30-NOV-2000; 2000US-00728422.
 XX (HYSE-) HYSEQ INC.
 PA Tang YT, Liu C, Drmanac RT, Asundi V, Zhou P, Xu C, Cao Y;
 PI Ma Y, Zhao QH, Wang D, Wang J, Zhang J, Ren F, Chen R, Wang ZW;
 PI Xue AJ, Yang Y, Wejhrman T, Goodrich R;
 XX WPI: 2001-476283/51.
 XX N-PSDB; AAK52874.
 PT Nucleic acids encoding polypeptides with cytokine-like activities, useful
 PT in diagnosis and gene therapy.
 XX Claim 20; Page 319; 6221pp; English.
 PS The invention relates to polynucleotides (AAK51456-AAK53435) and the
 CC encoded polypeptides (AAM78323-AAK80302) that exhibit activity relating to
 CC cytokine, cell proliferation or cell differentiation or which may induce
 CC production of other cytokines in other cell populations. The
 CC polynucleotides and polypeptides are useful in gene therapy, vaccines or
 CC peptide therapy. The polypeptides have various cytokine-like activities,
 CC e.g. stem cell growth factor activity, haematopoiesis regulating
 CC activity, tissue growth factor activity, immunomodulatory activity and
 CC activin/inhibin activity and may be useful in the diagnosis and/or
 CC treatment of cancer, leukaemia, nervous system disorders, arthritis and
 CC inflammation. Note: Records for SEQ ID NO 2110 (AAK52581), 2111
 CC (AAK52582) and 3666 (AAM80020) are omitted as the relevant pages from the
 CC sequence listing were missing at the time of publication

XX SQ Sequence 510 AA;
 Query Match 86.0%; Score 2324; DB 4; Length 510;
 Best local Similarity 99.8%; Pred. No. 1.66-136;
 Matches 459; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
 QY 1 KSSPGQPEAGPEGAQRRPSQAAPAVEAGPGSSQAAPRKPEGAQAARTASGALRDVSELS 60
 DB 28 KSSPGQPEAGPEGAQRRPSQAAPAVEAGPGSSQAAPRKPEGAQAARTASGALRDVSELS 87
 QY 61 RQLEDILSTYCVNNQGGPGEDGAQCEPAPEPAEKSRITYVAENGPEPTPVVYGEKPS 120
 DB 88 RQLEDILSTYCVNNQGGPGEDGAQCEPAPEPAEKSRITYVAENGPEPTPVVYGEKPS 147
 QY 121 KQDPNTEIRQSDDEVGDHRRRPOEKKKAKGLGKITLLMQTLNTLSTPEKLAALCKKY 180
 DB 148 KQDPNTEIRQSDDEVGDHRRRPOEKKKAKGLGKITLLMQTLNTLSTPEKLAALCKKY 207
 QY 181 AELLEHRNSQKMKLLQKQSQLVQEKHLSREHSAVLARSKLSLCELOHNRSLK 240
 DB 208 AELLEHRNSQKMKLLQKQSQLVQEKHLSREHSAVLARSKLSLCELOHNRSLK 267
 QY 241 EGVQVAREEERKEVTSHFQVTLNDIQLQMHNRNSKLRQENNELAERLKKLIEQY 300
 DB 268 EGVQVAREEERKEVTSHFQVTLNDIQLQMHNRNSKLRQENNELAERLKKLIEQY 327
 QY 301 ELREHIDKVFHKDLQQLVDKLAQQAQEMKAEABERHOREKDFLLKEAVESQRMCELM 360
 DB 328 ELREHIDKVFHKDLQQLVDKLAQQAQEMKAEABERHOREKDFLLKEAVESQRMCELM 387
 QY 361 KQETHLKQQLALYTEKPEFQNTLSKSSSEVFTTFKQEMERQTKIKKLEKETTMYSRW 420
 DB 388 KQETHLKQQLALYTEKPEFQNTLSKSSSEVFTTFKQEMERQTKIKKLEKETTMYSRW 447
 QY 421 ESSNKALLEMAEBKTVRDKEGLQVKIQRLEKLCRALQT 460
 DB 448 ESSNKALLEMAEBKTVRDKEGLQVKIQRLEKLCRALQT 487

RESULT 11
 ABO07223
 ID ABO07223 standard; protein; 356 AA.
 XX ABO07223;
 XX 13-AUG-2003 (first entry)
 DT Human p53 modifying protein, SEQ ID 183.
 DE Human; p53 modifier; cytostatic; cancer; cytostatic; antiangiogenic;
 KW antiapoptotic; p53 pathway; breast cancer; colon cancer; kidney cancer;
 KW lung cancer; ovarian cancer; angiogenesis; cell cycle;
 KW apoptotic disorder; cell proliferation disorder.
 OS Homo sapiens.
 QY WO200299122-A1.
 FN 12-DEC-2002.
 PD 03-JUN-2002; 2002WO-US017382.
 XX 05-JUN-2001; 2001US-0296076P.
 XX 10-OCT-2001; 2001US-0328605P.
 XX 15-FEB-2002; 2002US-0357253P.
 XX (EXEL-) EXELIXIS INC.
 PA Friedman L, Plowman GD, Belvin M, Francis-Lang H, Li D, Funke RP;
 PI WPI: 2003-156859/15.
 XX N-PSDB; ACTD1344B.
 DR

Query Match	66.7%	Score 1801;	DB 6;	Length 356;
Best Local Similarity	100.0%;	Pred. No. 3.7e-104;		
Matches 356;	Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0;
QY	175	ALCKCYAELLLEHRNSQKQMKLLQKQSQQLVQSKHLRGHSHKAVLARSKGLSCLREQLR	234	
DB	1	ALCKCYAELLLEHRNSQKQMKLLQKQSQQLVQSKHLRGHSHKAVLARSKGLSCLREQLR	60	
QY	235	HNRSLSKEEGVQVAREEBEKKKEVTSHFQVTLNDIQLQMEQHNRNSKLRQNNMELAERLX	294	
DB	61	HNRSLSKEEGVQVAREEBEKKKEVTSHFQVTLNDIQLQMEQHNRNSKLRQNNMELAERLX	120	
QY	295	KLIQEQYELREBZHIDKVPKHLQLOQLVDAKLQQAQEWMLKEAEERHOREKDFILKEAVESQ	354	
DB	121	KLIQEQYELREBZHIDKVPKHLQLOQLVDAKLQQAQEWMLKEAEERHOREKDFILKEAVESQ	180	
QY	355	RMCELMKQOETHLKQOLALYETKPEPFQNTLSKSSSEVFTTFKQEMERMTKKIKKLKSKETT	414	
DB	181	RMCELMKQOETHLKQOLALYETKPEPFQNTLSKSSSEVFTTFKQEMERMTKKIKKLKSKETT	240	
QY	415	MYRSRWESSNKALLEWAEKEETVDKELEGLOVKIQRLEKLCRALQYTERNDLNKRVQDLSA	474	
DB	241	MYRSRWESSNKALLEWAEKEETVDKELEGLOVKIQRLEKLCRALQYTERNDLNKRVQDLSA	300	
QY	475	GGQGSITDGGPERPDPGPGQAQAPSSPRVTAPCYPGAPSPSTEASGGTGPOBPTTSARA	530	
DB	301	GGQGSITDGGPERPDPGPGQAQAPSSPRVTAPCYPGAPSPSTEASGGTGPOBPTTSARA	356	

RESULT 12
AAW78757
ID AAW78757 standard; protein; 358 AA.
XX

AA	Sequence	358 AA;	Query Match	64.9%;	Score	1754.5;	DB	4;	Length	358;
Best	Local Similarity	98.3%;	Pred. No.	2.9e-101;						
Matches	349;	Conservative	0;	Mismatches	5;	Indels	1;	Gaps	1;	
QY	177	CKKY-ABLLBEHRNSQKMKLKQKQSOLVQKHILRGHSHKAVILARSKLSLCELORH	235							
DB	4	CPFPFCPLBEHRNSQKMKLKQKQSOLVQKHILRGHSHKAVILARSKLSLCELORH	63							
QY	236	NRSIKKEGVQARBEERKEKKEVTSHTPQVTLNDTLQMEQHNRNSKLAQENNELAERLKK	295							
DB	64	NRSIKKEGVQARBEERKEKKEVTSHTPQVTLNDTLQMEQHNRNSKLAQENNELAERLKK	123							
QY	296	LIEQVEILREBHIDKVFPHKDLQOQLVDAKLQQAQEMLKAEABERHOREKDFLLKXAVESQR	355							
DB	124	LIEQVEILREBHIDKVFPHKDLQOQLVDAKLQQAQEMLKAEABERHOREKDFLLKXAVESQR	183							

XX (EXEL-) EXELIXIS INC.
PA Friedland L, Plowman GD, Belvin M, Francis-Lang H, Li D, Funke RP;
PI WPI; 2003-156859/15.
DR N-PSDB; ACD13396.
XX Identifying modulators of the p53 pathway for use in treating apoptotic
PT or cell proliferation disorders, comprises screening for agents that
PT modulate activity of a human ortholog of genes that modify the p53
PT pathway in Drosophila.
XX Example 2; Page 532-533; 678pp; English.
PS
XX The invention relates to identifying (M1) a candidate p53 pathway
XX modulating agent, by contacting an assay system comprising a purified HM
CC polypeptide [human orthologue of genes that modify the p53 pathway in
CC Drosophila] or nucleic acid with a test agent under conditions, where but
CC for the presence of the test agent, the system provides a reference
CC activity, and detecting a test agent-biased activity of the assay system.
CC Also included are modulating (M2) a p53 pathway of a cell (comprising
CC contacting a cell defective in p53 function with a candidate modulator
CC that specifically binds to a HM polypeptide comprising an HM amino acid
CC sequence, where p53 function is restored), modulating (M3) a p53 pathway
CC in a mammalian cell (comprising contacting the cell with an agent that
CC specifically binds an HM polypeptide or nucleic acid) and diagnosing (M4)
CC a disease in a patient (comprising: (a) obtaining a biological sample
CC from the patient; (b) contacting the sample with a probe for HM
CC expression; (c) comparing the results with a control; and (d) determining
CC whether the comparison indicates a likelihood disease). (M1) is useful
CC for identifying modulators of the p53 pathway. A probe for HM expression
CC is useful for diagnosing breast, colon, kidney, lung and ovarian cancer,
CC in a patient, where the cancer has greater than 25 % expression level.
CC Modulators identified by (M1) are useful in a variety of diagnostic and
CC therapeutic applications, where disease or disorder prognosis is related
CC to defects in the p53 pathway, such as, angiogenesis, apoptotic or cell
CC proliferation disorders (e.g. cancer). Another two new methods (M2 and
CC M3) are useful for modulating the p53 pathway of a cell, thus restoring
CC the p53 function of the cell, so that the cell undergoes normal
CC proliferation or progression through the cell cycle. (M2) and (M3) are
CC also useful for treating defects in the p53 pathway such as angiogenic,
CC apoptotic or cell proliferation disorders. The present sequence
CC represents a human p53 pathway modifying protein
XX
XX Sequence 395 AA;
Query Match 46.8%; Score 1265.5; DB 6; Length 395;
Best Local Similarity 65.2%; Pred. No. 9e-71;
Matches 255; Conservative 59; Mismatches 50; Indels 27; Gaps 5;
QY 153 GKETLLMOTLNTLSTPEKLAALCKKYAELEHRSNOKMLQKQKQSLVQEKDHLR 212
DB 3 GKVELLMQALNTLSTPEKLAALCKKYADLSESRVQKMLQKQQAIVKEKVLQ 62
QY 213 GEHSKAVLARSKLESCLRELOHNRSLKEGVQARBEERKEVTSFHVTLNDLQLM 272
DB 63 SEHSKAVLARSKLESCLRELOHNRSLKEGVQARBEERKEVTSFHVTLNDLQLM 122
QY 273 FOHNRSLKROENNELARLKLIEOYELREHIDKVFKHDLQOOLVDKQLQQAQEM 332
DB 123 EOHDIRHAKURQENIEGKLLKIEQYALREHIDKVFKEHLOOQVLVDKQLQQTOLI 182
QY 333 KEAEERHOREKDFLLKEAVESQRMCELMKQOETHLKOQALYTEKPEFQNTLSKSEVP 392
DB 183 KEAEERHOREKDFLLKEAVESQRMCELMKQOETHLKOQALYTEKPEFQNTLSKSEVP 242
QY 393 TTFQOEMTKTKIKLEKETMYRSWESSNKALLEWAEKTVRDKELGLOVKIQRLE 452
DB 243 TTFQOEMTKTKIKLEKETMYRSWESSNKALLEWAEKTVRDKELGLOVKIQRLE 302
QY 453 KLCLALOTERNELNKRVO-----DLSAGGQGSLLT--DSGPE-----RR 488
KLCLALOTERNELNKRVEVLKQSVISIAAIAKANRDLATPWPQCTALDSKELNTSSKR 362
489 PEGPCAQA-PSSPR--VTEAPCYEGAPSTEA 516
363 ALGAHLEAPKQSAVQKPPSTGSAFAIRS 393

DB 303 KLCLALOTERNELNKRVEVLKQSVISIAAIAKANRDLATPWPQCTALDSKELNTSSKR 362
QY 489 PEGPCAQA-PSSPR--VTEAPCYEGAPSTEA 516
DB 363 ALGAHLEAPKQSAVQKPPSTGSAFAIRS 393
RESULT 15
ABR43284
ID ABR43284 standard; protein; 684 AA.
AC ABR43284;
XX 09-JUL-2003 (first entry)
XX Human neurotrophin-associated protein NTRAN-14 SEQ ID NO:14.
XX Human; neurotrophin-associated protein; NTRAN; cytostatic; anti-HIV;
KW antiarteriosclerotic; anticonvulsant; neurotrophic; neuroprotective; stroke;
KW cerebroprotective; antiallergic; antiinflammatory; thyromimetic; cancer;
KW antidiabetic; gene therapy; cell proliferative disorder; atherosclerosis;
KW neurological disorder; epilepsy; Huntington's disease; immune disorder;
KW inflammatory disorder; AIDS; allergy; developmental disorder; diabetes;
KW hypothyroidism; Cushing's syndrome; endocrine disorder; infection.
OS Homo sapiens.
XX WO2003025129-A2.
XX 27-MAR-2003.
XX 12-SEP-2002; 2002WO-US029219.
XX 14-SEP-2001; 2001US-0322180P.
XX 28-SEP-2001; 2001US-0326096P.
XX 04-OCT-2001; 2001US-0327448P.
XX 26-OCT-2001; 2001US-0345837P.
XX 02-NOV-2001; 2001US-0343903P.
XX 27-NOV-2001; 2001US-0344020P.
XX 07-DEC-2001; 2002US-0340266P.
XX 04-JAN-2002; 2002US-0345008P.
XX 18-MAR-2002; 2002US-0365645P.
XX 10-MAY-2002; 2002US-0379887P.
XX (INCY-) INCYTE GENOMICS INC.
XX Honchell CD, Warren BA, Borowsky ML, Griffin JA, Li JX, Lee SY;
PI Yue H, Forsythe IJ, Marquis JP, Gietzen KJ, Baughn MR, Tran UX;
PI Lehr-Mason PM, Tang YT, Ramkumar J, Emerling BM, Lee EA, Elliott VS;
PI Hafalia AJA, Duggan BM, Kawla NK, Kable AE, Chang H, Khare R;
PI Becha SD, Jin P, Lee S;
XX WPI; 2003-363137/34.
XX N-PSDB; ACC68992.
XX New human neurotrophin associated proteins (NTRAN), useful for
PT diagnosing, treating and preventing diseases or conditions associated
PT with the aberrant NTRAN expression e.g. cancer, AIDS, diabetes, epilepsy,
PT or infections.
XX Claim 1; Page 201-203; 240pp; English.
XX ACC68992 to ACC69003 encode the human neurotrophin-associated
CC proteins given in ABR43271 to ABR43295, designated NTRAN-1 to NTRAN-25
CC (i). (i) have cytostatic, antiarteriosclerotic, anticonvulsant,
CC neurotrophic, neuroprotective, cerebroprotective, anti-HSV, antiallergic,
CC antiinflammatory, thyromimetic and antidiabetic activities, and can be
CC used in gene therapy. The NTRAN polypeptides and polynucleotides are
CC useful in diagnosing, treating and preventing diseases or conditions
CC associated with the decreased expression or overexpression of NTRAN, such
CC as cell proliferative (e.g. cancer, atherosclerosis), neurological (e.g.
CC epilepsy, Huntington's disease, stroke), immune/inflammatory (e.g. AIDS,
CC allergies), developmental (e.g. hypothyroidism, Cushing's syndrome) or

CC endocrine (e.g. diabetes) disorders, or infections. They are also useful
CC in assessing the effects of exogenous compounds on the expression of
CC nucleic acid and amino acid sequences of NTRAN. The NTRANs or their
CC fragments are useful in screening compounds for effectiveness as agonist
CC or antagonist of the polypeptides, or in altering the expression of the
CC target polynucleotide and compounds that specifically bind to or modulate
CC the activity of the polypeptide

XX Sequence 684 AA;

Query Match 46.7%; Score 1260.5; DB 6; Length 684;
Best Local Similarity 46.3%; Pred. No. 3.5e-70;
Matches 272; Conservative 91; Mismatches 135; Indels 89; Gaps 9;
QY 4 PQOEPAGP--EQAQEPSCAAPAVEAEGPCSQAPKPEGAQAQTAQSGALRDVSEELSR 61
DB 18 PGSSSLPHNGLEKEDGQ-----DSPTVPQPEK-----BASVHPDISEELNR 61
QY 62 QLEDILSTYCVNNGGPGEDGAGQAPPAEPDAEKRTTVARNGBEPPTVPVYGEKPSK 121
DB 62 QLEDIINTY--GSAASTAGKGSARASEQENAESPDN---EDGCEETTEAGREPVAS 116
QY 122 GPNTPEIRIQSDVGRDHRHRRPQKKAGLKGKTEITLLMOTLNTLSTPEKLAALCKKYA 181
DB 117 GPPTVTK-----EPVSNKEQK--LEKILKGLGKEANLLMNLKLTPEKEDPLFKKYA 170
QY 182 ELLEHRSQKQMLLQKQSQVQKDLHRLGHSKAVLARSKLSLCSLRELOHNRSLKE 241
DB 171 ELLDEHRTQKLLKLLQKQVQIQEKDQLQGEHSRAILARSKLSLCSLRELOHNRSLKE 230
QY 242 EGVORAREEERKKEVTSFQVTLNDILOMEQHNRNSKLRQENMELAEKRLKGLBOYE 301
DB 231 EALQAREEERKKEVTSFQVTLNDILOMEQHNRNSKLRQENMELAEKRLKGLBOYE 290
QY 302 LREEHIDKVFHKDLQQLVDAKLOQAQEMLKEAEERHOREKDFLLKEAVESQRMCELA 361
DB 291 LREELDKIFKHRLQQLVDAKLEQAQEMMKEAEERHOREKDFLLQNAEAWKLOAKVLK 350
QY 362 QQETHLQQLALYTEKPEEFQNTLSKSEVFTTFKQEMKMTKIKKLEKETTMYR 421
DB 351 EQETVLQQLTLYSGRFEFQNTLSKSEVFTTFKQEMKMTKIKKLEKETTMYR 410
QY 422 SNNKALLEMAEKTVDKLEGLQVKIQRLEKLCRALQTERNDLNKRVQDLSAGGQSIT 481
DB 411 NCKALLDMIEEALRAKEVCFVMKIGRLNLCRALQERNEHLKIKRDASISEKDDQS 470
QY 482 DSGPERRPEG----- 491
DB 471 QHNSDEEPSNVSVDQEDAEVNSVQTAVYNLATAPMIHHPESTPHQSKETQPEIGSS 530
QY 492 -----PGAQAPSSP-----RVTEAPCVPGAPSTASGQTGPQSPSTA 528
DB 531 QESADAALKEPQPLIPSDSESPLPLTPQAEAGGSDAEPSPKA 577

RESULT 16

AAV17863

ID AAV17863 standard; protein; 676 AA.

XX AAV17863;

XX 16-AUG-1999 (first entry)

XX Neurite extending activity protein.

XX Neurite extending activity; anti-dementia; memory; brain function;

XX Dementia.

XX Mus sp.

XX JP11147897-A.

XX 02-JUN-1999.

XX

XX

XX

XX

XX

XX

XX 13-NOV-1997; 97JP-00331242.
XX 13-NOV-1997; 97JP-00331242.
XX (AGEN) AGENCY OF IND SCI & TECHNOLOGY.
XX WPI; 1999-379889/32.
XX N-PSDB; AAX80156.
XX New protein with neurite extending activity - useful for treating
XX dementia.
XX Claim 1; Page 5-8; 9pp; Japanese.
XX The present sequence represents a protein (I) which has neurite extending
XX activity. (I) may be administered to patients to prevent dementia or to
XX improve memory and brain function. (I) is also used as a neurite-
XX extending agent
XX Sequence 676 AA;
Query Match 45.6%; Score 1233; DB 2; Length 676;
Best Local Similarity 53.9%; Pred. No. 1.8e-68;
Matches 255; Conservative 84; Mismatches 100; Indels 34; Gaps 7;
QY 2 SSPQEPAGPAGAPERPSCAAPAVEAEGPCSQAPKPEGAQAQTAQSGALRDVSEELSR 61
DB 15 TSPTQDNQG-----QSKAEPVPVPSQLSP-TNQTSAQPEMATC-----DISEELNR 59
QY 62 QLEDILSTYCVNNGGPGEDGQ-----QCEPAEPDAEKRTTVARNGBEPPTVPVYGEK 117
DB 60 QLEDIITY-----GSAASLVEKEGTTAETDKPEKEDVGSMEDEACEDVNESEKD 110
QY 118 BPSKDPNTEIRIQSDVGRDHRHRRPQKKAGLKGKTEITLLMOTLNTLSTPEKLAALC 177
DB 111 KPAPGDAS-----RAKPSASKEQK--LEKILKGLGKEATLLMQSLNKLTPPEKLDLLF 164
QY 178 KYAELLEHRNSQKMLLQKQSQVQKDLHRLGHSKAVLARSKLSLCSLRELOHNR 237
DB 165 KYAELLEHRABEQQLKYLQKQVQIQEKDQLQGEHSRAILARSKLSLCSLRELOHNR 224
QY 238 SLKEEGVORAREEERKKEVTSFQVTLNDILOMEQHNRNSKLRQENMELAEKRLKLI 297
DB 235 TLKEETIQAREEERKKEVTSFQVTLNDILOMEQHNRNSKLRQENMELAEKRLKLI 284
QY 298 BOYELREHIDKVFHKDLQQLVDAKLOQAQEMLKEAEERHOREKDFLLKEAVESQRM 357
DB 285 BOYELREHIDKVFHKDLQQLVDAKLOQAQEMLKEAEERHOREKDFLLKEAVESQRM 344
QY 358 ELKQOETHLQQLALYTEKPEEFQNTLSKSEVFTTFKQEMKMTKIKKLEKETTMYR 417
DB 345 XWLKQOETHLQQLALYTEKPEEFQNTLSKSEVFTTFKQEMKMTKIKKLEKETTMYR 404
QY 418 SRWESSNKALLEMAEKTVDKLEGLQVKIQRLEKLCRALQTERNDLNKRVQ 470
DB 405 SRFENCNRLALDMTSEKAMRTKVEYCFVLKIQRLEKLCRALQERNEHLKIK 457
RESULT 17
AAW37883
ID AAW37883 standard; protein; 386 AA.
XX AAW37883;
XX 28-AUG-1998 (first entry)
XX BRC1 modulator protein 091-132Q20.
XX BRC1 modulator protein; 091-132Q20; breast cancer antigen 1;
XX tumour suppressor protein; diagnosis; therapy; human.
XX Homo sapiens.
XX OS


```
XX Key Location/Qualifiers
FH Domain 124..143
FT /note="leucine zipper motif"
XX WC9810066-A1.
XX 12-MAR-1998.
XX 06-AUG-1997; 97WO-US013944.
XX 04-SEP-1996; 96US-0025601P.
XX (ONVX-) ONYX PHARM INC.
XX Rubinfield B, Polakis P, Ligenfelter C, Vuong TT;
XX WPI; 1998-193616/17.
XX N-PSDB; AAV29064.
XX Breast cancer antigen 1 modulator protein - useful for diagnosing
XX diseases involving unwanted cell growth, e.g. breast cancer, and for
XX producing therapeutics for treatment of such diseases.
XX Example 1; Fig 3; 73pp; English.
XX This polypeptide comprises a 46 kDa BRCA1 modulator protein that binds to
XX the tumour suppressor gene product BRCA1, and which is characterised by a
XX leucine zipper motif. Its amino acid sequence was deduced from the
XX nucleotide sequence of a cDNA clone (see AAV29064), designated 031-132Q20
XX (ATCC 98143), isolated from a HeLa cell cDNA library using a yeast two-
XX hybrid assay. 3 cDNA clones (see also AAV29062 and AAV29063) coding for
XX BRCA1 modulator proteins (see AAV37881-83) have been characterised.
XX Vectors and host cells comprising the isolated nucleic acid sequences are
XX claimed, as well as a process for producing BRCA1 modulator protein by
XX culturing these host cells. BRCA1 modulator proteins and nucleic acids
XX can be used to diagnose diseases involving unwanted cell growth, e.g.
XX breast cancer, and to identify compounds that alter BRCA1 interaction
XX with BRCA1 modulators for the treatment of such diseases
XX SQ Sequence 386 AA;
XX Query Match 38.4%; Score 1038.5; DB 2; Length 386;
XX Best Local Similarity 56.0%; Pred. No. 1.2e-56;
XX Matches 225; Conservative 55; Mismatches 99; Indels 23; Gaps 6;
XX 13 GAQRPSPQAPAVAEAGPGSSQAPRK--PEGAQARTAQSGALRDVSEELSRQLEDILSTY 70
XX 4 GAEE-----ATEAGRGRRRSFRQKFEIGTMEAGICGLGVKADMLCNSQSNLILQHQ 56
XX 71 CVDNNGGPGEDGAGQEPAPEDAEKSRITYVARGSEPTPVYVGEKPSKGDPTNTEIR 130
XX 57 --GSGCGGTSNKHSLEEDSGSDFTENRLVS-----PAYCTQESREIPGG----EAR 104
XX 131 QSDVGVGRDRHRRPOEKKAKGLGKEITLLMOTLNTLSTPEKLAALCKYAELEHRNS 190
XX 105 TDPDGGQDSECNKKEKT--LGKEVILLMQALNTLSTPEKLAALCKYADLLESRSV 162
XX 191 QKQMLKQKQSQQLVQEKHURGHGSKAVILARSKLSLCRELQHRNLSLKEGVQVQAREE 250
XX 163 QKQMLKQKQQAIVQEKVHLQSEHSKAILARSKLSLCRELQHRNLTKEENMQQAREE 222
XX 251 EEKKEKVTSHFQVTLNDIQLQMOHNRNSKLRQENMELAEKLLIQEYELREHIDKV 310
XX 223 EERIEATAHQITLNEIQAQEQBDIHNAKLRENIELGEKLLIQEYALREHIDKV 282
XX 311 FAKHDLQQLVDALQQAQEMKLABERHQBKDFLLKGAVESORMCELMKQOETHLKQQ 370
XX 283 FKHRELQQLVDALQQTTLQKEADKHQREERFLKZATESRHKYEQNKQEQVQLKQQ 342
XX 371 LALYTEREEFQNTLSKSSVEFTTFKQEMQNTKKIKKLEKE 412
XX 343 LSLYNDKEEFQTTWAKSNELFTTFQEMEQNTKKIKKQCKK 384
```

```
RESULT 18
AAV30151
ID AAV30151 standard; protein; 386 AA.
XX AC AAV30151;
XX 27-OCT-1999 (first entry)
XX Amino acid sequence of a BRCA1 modulator protein.
XX Modulator protein; BRCA1; tumour suppressor protein; breast cancer;
XX ovarian cancer; cell growth; cell proliferation.
XX Homo sapiens.
XX Key Location/Qualifiers
FH Region 125..143
FT /note="leucine zipper motif"
XX US5948643-A.
XX 07-SEP-1999.
XX 13-AUG-1997; 97US-00968751.
XX 13-AUG-1997; 97US-00968751.
XX (ONVX-) ONYX PHARM INC.
XX Rubinfield B, Ligenfelter C, Vuong TT, Polakis PG;
XX WPI; 1999-517952/43.
XX N-PSDB; AAX86756.
XX Modulator proteins that bind to and modulate the activity of the BRCA1
XX tumor suppressor gene product, useful for the treatment of ovarian and
XX breast cancer.
XX Example 1; Fig 3; 35pp; English.
XX The present sequence represents a modulator protein, that binds to and
XX modulate the activity of the BRCA1 gene product (BRCA1). The BRCA1
XX protein has been characterized as a tumour suppressor protein.
XX Alterations in the amino acid sequence of BRCA1 causes breast and ovarian
XX cancers by removing the controls on cell growth and proliferation.
XX Research has shown that different regions on the BRCA1 molecule have
XX different effects on cell growth and tumour suppression (e.g. full length
XX truncated BRCA1 has no effect on breast cancer cell growth but will
XX inhibit ovarian cancer cell growth). It has been suggested that different
XX host cell factors (e.g. proteins) interact with different regions of the
XX BRCA1 to control its function. The identification of these proteins (e.g.
XX BRCA1MP) will facilitate the development of novel diagnostic methods and
XX new therapeutics for identifying and treating cancers caused by changes
XX in the expression or activity of BRCA1
XX SQ Sequence 386 AA;
```

```
Query Match 38.4%; Score 1038.5; DB 2; Length 386;
Best Local Similarity 56.0%; Pred. No. 1.2e-56;
Matches 225; Conservative 55; Mismatches 99; Indels 23; Gaps 6;
QY 13 GAQRPSPQAPAVAEAGPGSSQAPRK--PEGAQARTAQSGALRDVSEELSRQLEDILSTY 70
DB 4 GAEE-----ATEAGRGRRRSFRQKFEIGTMEAGICGLGVKADMLCNSQSNLILQHQ 56
QY 71 CVDNNGGPGEDGAGQEPAPEDAEKSRITYVARGSEPTPVYVGEKPSKGDPTNTEIR 130
DB 57 --GSGCGGTSNKHSLEEDSGSDFTENRLVS-----PAYCTQESREIPGG----EAR 104
QY 131 QSDVGVGRDRHRRPOEKKAKGLGKEITLLMOTLNTLSTPEKLAALCKYAELEHRNS 190
```


Db 105 TDPDQDSECNNEKT--LGKEVLLMQALNTLSTPEKLAALCKKYADLLSESRV 162
QY 191 QKMKLLQKKQOLVOEKDHLRGHSHKAVIARSKLESCLREIQRNRSUKRGVQARRE 250
Db 163 QKMKILQKKQAOIVKEVHLOSEHSAKILASKLESCLREIQRNRSUKRGVQARRE 222
QY 251 ESKRKEVTSFQVTLNDIOLQMEHNRNSKLRQENMELAEKLIQOYELREHIDKV 310
Db 223 ERRIETATAFQITLNEIQAEQDHIHNAKLRQENIELGCKLKLIEQYALREHIDKV 282
QY 311 FKHKDQQLVDAKQQAQOEMLKEABERHOREKDFLLKEAVESORMCELMKQOETHLQK 370
Db 283 FKHEKQQLVDAKQQAQOETHLKEABERHOREKDFLLKEAVESORMCELMKQOETHLQK 342
QY 371 LALYTEKPEFQVTSKSEVFTTPKQEMKTKIKKLEKE 412
Db 343 LSLYMDKPEFQVTSKSEVFTTPKQEMKTKIKKLEKE 384

RESULT 19
AAB63259
ID AAB63259 standard; protein; 204 AA.

XX AAB63259;
XX 26-MAR-2001 (first entry)
XX Human breast cancer associated antigen protein sequence SEQ ID NO:621.
DE Human; breast cancer; gastric cancer; prostate cancer; diagnosis;
XX cancer associated antigen; cytostatic; cancer vaccine.
XX Homo sapiens.

OS WO200073801-A2.
PN 07-DEC-2000.
XX 26-MAY-2000; 2000WO-US014749.
XX 28-MAY-1999; 99US-0136526P.
PR 10-SEP-1999; 99US-0153454P.

XX (LUDW-) LUDWIG INST CANCER RES.
PA Obata Y;
XX WPI; 2001-025274/03.
XX Nucleic acids encoding breast, gastric and prostate cancer associated
PT antigen precursors, useful for diagnosing and treating a condition
PT characterized by expression of an abnormal amount of a protein, e.g.
XX cancer.

XX Example 1; Page 486-487; 799pp; English.
XX AAF22422 to AAF22626, AAF22627 to AAF22773 and AAF22774 to AAF23014
CC represent nucleotide sequences encoding human breast, gastric and
CC prostate cancer associated antigen precursors (CAAP) respectively.
CC AAB63232 to AAB63467, AAB63468 to AAB63721 and AAB63722 to AAB63970
CC represent human breast, gastric and prostate CAAP protein sequence
CC respectively. CAAPs have cytostatic activity and can be used in the
CC reduction of cancer vaccines. The human CAAP proteins, peptides, nucleic
CC acids or anti-CAAP antibodies are useful for diagnosing and treating a
CC condition characterised by expression of an abnormal amount of a protein,
CC e.g. cancer
XX Sequence 204 AA;

Query Match 30.2%; Score 815; DB 4; Length 204;
Best Local Similarity 76.0%; Pred. No. 4.5e-43;
Matches 155; Conservative 30; Mismatches 19; Indels 0; Gaps 0;

QY 163 INTLSTPEKLAALCKKYAELEHNRNSQKMKLLQKKQOLVOEKDHLRGHSHKAVLAR 222
Db 1 INTLSTPEKLAALCKKYADLLSESRVQKMKILQKKQAOIVKEVHLOSEHSAKILAR 60
QY 223 SKLESCLREIQRNRSUKRGVQARREKKEVTSFQVTLNDIOLQMEHNRNSKL 282
Db 61 SKLESCLREIQRNRSUKRGVQARREKKEVTSFQVTLNDIOLQMEHNRNSKL 120
QY 283 ROENMELAEKLIQOYELREHIDKVFKHDKQQLVDAKQQAQOEMLKEABERHORE 342
Db 121 ROENIELGCKLKLIEQYALREHIDKVFKHDKQQLVDAKQQAQOEMLKEABERHORE 180
QY 343 KDFLLKEAVESORMCELMKQOETH 366
Db 181 REFLLKEAVESORMCELMKQOETH 204

RESULT 20
AAO08301
ID AAO08301 standard; protein; 475 AA.

XX AAO08301;
XX 06-NOV-2001 (first entry)
XX Human polypeptide SEQ ID NO 22193.

XX Human; cytokine; cell proliferation; cell differentiation; gene therapy;
KW vaccine; peptide therapy; stem cell growth factor; haematopoiesis;
KW tissue growth factor; immunomodulatory; cancer; leukaemia;
KW nervous system disorders; arthritis; inflammation.

OS Homo sapiens.

XX WO200164835-A2.

XX 07-SEP-2001.

XX 26-FEB-2001; 2001WO-US004927.

XX 28-FEB-2000; 2000US-00515126.

XX 18-MAY-2000; 2000US-00577409.

XX (HYSE-) HYSEQ INC.

XX Tang YT, Liu C, Drmanac RT;

XX WPI; 2001-514838/56.

XX N-PSDB; AAI88322.

XX Isolated nucleic acids and polypeptides, useful for preventing diagnosing

PT and treating e.g. leukemia, inflammation and immune disorders.

XX Claim 20; SEQ ID NO 22193; 1399pp + Sequence Listing; English.

XX The invention relates to human polynucleotides (AAI79941-AAI93841) and
CC the encoded proteins (AAO0010-AAO13910) that exhibit activity elating to
CC cytokine, cell proliferation or cell differentiation or which may induce
CC production of other cytokines in other cell populations. The
CC polynucleotides and polypeptides are useful in gene therapy, vaccines or
CC peptide therapy. The polypeptides have various cytokine-like activities,
CC e.g. stem cell growth factor activity, haematopoiesis regulating
CC activity, tissue growth factor activity, immunomodulatory activity and
CC activin/inhibin activity and may be useful in the diagnosis and/or
CC treatment of cancer, leukaemia, nervous system disorders, arthritis and
CC inflammation. Note: The sequence data for this patent did not form part
CC of the printed specification, but was obtained in electronic format
CC directly from WIPO at ftp.wipo.int/pub/published_pct_sequences

XX Sequence 475 AA;

Query Match 30.0%; Score 810; DB 4; Length 475;
Best Local Similarity 47.7%; Pred. No. 2.5e-42;

```
Matches 166; Conservative 54; Mismatches 68; Indels 60; Gaps 2;
QY 241 EGVQARAEERKEVTSHFQVTLNDIQLQMEQHNRNKLQENMELAEKLIQY 300
Db 29 BEALQARAEERKEVTSHFQVTLNDIQLQMEQHNRNKLQENMELAEKLIQY 88
QY 301 ELREEHIDKVPKDKLQOOLVDKLAQAEMLKEAEERHOREKDFLLKEAVESQRMCELM 360
Db 89 ELREEHIDKVPKDKLQOOLVDKLAQAEMLKEAEERHOREKDFLLKEAVESQRMCELM 148
QY 361 KOQETHLKQALALYTERKEFEFQNTLSKSSSEVFTTFKQEMERMTKKIKLEKETTWYRSRW 420
Db 149 KEQETVLQALTLYSGRFEERFQSTLTLSKSNVFAITFKQEMDNTTKMKKLEKDTATWKARF 208
QY 421 ESSNKALLEMAEKTVRDKLEGLQVKIQRLEKLCRALQTERNDLNKRVQDLSAGGQSL 480
Db 209 ENCNKALLDMIEKALRAKRYECFVYMKIGRLNLCRALQBERNELHKKIRDABISEKDDQ 268
QY 481 TDSGPRRPEG----- 491
Db 269 SQHNSDEPESNVSDQIDAEVNSVQTAVKVLATAPMIIHPSTPHQSKETQPEIGS 328
QY 492 -----PGAQAPSSP-----RVTEAPCTPGAPSTASGQTGPQSTSA 528
Db 329 SQESADAALKEPEQPPLIPSRDSESPFLPTPQAEAGGSDABEPSSKA 376

RESULT 21
ADC33216
ID ADC33216 standard; protein; 475 AA.
XX AC ADC33216;
XX DT 18-DEC-2003 (first entry)
XX DE Human novel contig-encoded polypeptide sequence, SEQ ID NO:3298.
XX KW Human; diagnostic; drug screening; forensics; gene mapping;
XX KW biodiversity assessment; Parkinson's disease; Alzheimer's disease;
XX KW neurodegenerative diseases; anaemia; platelet disorder; wound; burns;
XX KW ulcers; osteoporosis; autoimmune disease; cancer;
XX KW molecular weight marker; food supplement; antiparkinsonian; nootropic;
XX KW neuroprotective; antianaemic; anticoagulant; thrombolytic; vulnerary;
XX KW antitumor; osteopathic; immunosuppressive; antiinflammatory; cytostatic;
XX KW gene therapy; chromosome 6.
XX OS Homo sapiens.
XX PN W02003029271-A2.
XX PD 10-APR-2003.
XX PF 24-SEP-2002; 2002WO-US030474.
XX PR 24-SEP-2001; 2001US-0324631P.
XX PX (HYSE-) HYSEQ INC.
XX PY Tang TY, Zhang J, Ren F, Xue AJ, Zhao QA, Wang J, Wehrman T;
XX PZ Zhou P, Ghosh M, Wang D, Ma Y, Asundi V, Wang Z, Weng G;
XX PI Haley-Vicente D, Drmanac RT;
XX PP WPI: 2003-371981/35.
XX PR N-PSDB; ADC32449.
XX PT New polynucleotide and polypeptide useful for diagnosing, preventing or
XX PT treating conditions such as neurodegenerative diseases, anemias, platelet
XX PT disorders, wounds, burns, ulcers, osteoporosis, autoimmune diseases or
XX PT cancer.
XX PS Example 2; SEQ ID NO 3298; 1185pp; English.
XX DE The invention relates to 971 novel human cDNA sequences (ADC29919-
XX
```

```
CC ADC30899) and the polypeptides they encode (ADC30890-ADC31860). The
CC invention also relates to nucleic acid sequences over 99% identical with
CC the novel human cDNAs. The invention additionally encompasses expression
CC vectors and host cells comprising a nucleic acid of the invention; the
CC recombinant production of a polypeptide of the invention; an antibody
CC against a polypeptide of the invention; a method of detecting
CC polynucleotides or polypeptides of the invention; and methods of
CC identifying a compound which binds to a polypeptide of the invention. The
CC invention further discloses methods of preventing, treating or
CC ameliorating a medical condition; kits comprising polynucleotide probes
CC and/or monoclonal antibodies for carrying out the methods of the
CC invention; methods for the identification of compounds that modulate the
CC expression or activity of the polynucleotide and/or polypeptide; and 767
CC contig sequences corresponding to the cDNA sequences of the invention
CC (ADC31861-ADC32627) and the polypeptides encoded by the contigs (ADC32628
CC -ADC33394). The nucleic acids and polypeptides of the invention are
CC useful in diagnostics, drug screening, forensics, gene mapping, in the
CC identification of mutations responsible for genetic disorders or other
CC traits, for assessing biodiversity, and in producing many other types of
CC data and products dependent on DNA and amino acid sequences. They are
CC also used for treating degenerative diseases, anaemia, platelet
CC disorders, wounds, burns, ulcers, osteoporosis, autoimmune diseases or
CC cancer. The nucleic acids may also be used as hybridisation probes or
CC primers, and in the recombinant production of a protein. The polypeptides
CC are also useful in generating antibodies, as molecular weight markers,
CC and as food supplements. The present sequence represents a human contig-
CC encoded polypeptide sequence used in an example of the invention. Note:
CC The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX SQ Sequence 475 AA;
XX Query Match 30.0%; Score 810; DB 7; Length 475;
XX Best Local Similarity 47.7%; Pred. No. 2.5e-42;
XX Matches 166; Conservative 54; Mismatches 68; Indels 60; Gaps 2;
QY 241 EGVQARAEERKEVTSHFQVTLNDIQLQMEQHNRNKLQENMELAEKLIQY 300
Db 29 BEALQARAEERKEVTSHFQVTLNDIQLQMEQHNRNKLQENMELAEKLIQY 88
QY 301 ELREEHIDKVPKDKLQOOLVDKLAQAEMLKEAEERHOREKDFLLKEAVESQRMCELM 360
Db 89 ELREEHIDKVPKDKLQOOLVDKLAQAEMLKEAEERHOREKDFLLKEAVESQRMCELM 148
QY 361 KOQETHLKQALALYTERKEFEFQNTLSKSSSEVFTTFKQEMERMTKKIKLEKETTWYRSRW 420
Db 149 KEQETVLQALTLYSGRFEERFQSTLTLSKSNVFAITFKQEMDNTTKMKKLEKDTATWKARF 208
QY 421 ESSNKALLEMAEKTVRDKLEGLQVKIQRLEKLCRALQTERNDLNKRVQDLSAGGQSL 480
Db 209 ENCNKALLDMIEKALRAKRYECFVYMKIGRLNLCRALQBERNELHKKIRDABISEKDDQ 268
QY 481 TDSGPRRPEG----- 491
Db 269 SQHNSDEPESNVSDQIDAEVNSVQTAVKVLATAPMIIHPSTPHQSKETQPEIGS 328
QY 492 -----PGAQAPSSP-----RVTEAPCTPGAPSTASGQTGPQSTSA 528
Db 329 SQESADAALKEPEQPPLIPSRDSESPFLPTPQAEAGGSDABEPSSKA 376

RESULT 22
ADC31613
ID ADC31613 standard; protein; 416 AA.
XX AC ADC31613;
XX DT 18-DEC-2003 (first entry)
XX DE Human novel polypeptide sequence, SEQ ID NO:1695.
XX
```

Human; diagnostic; drug screening; forensics; gene mapping; biodiversity assessment; Parkinson's disease; Alzheimer's disease; neurodegenerative diseases; anaemia; platelet disorder; wound; burns; ulcers; osteoporosis; autoimmune disease; cancer; molecular weight marker; food supplement; antiparkinsonian; neurotropic; neuroprotective; antianemic; anticoagulant; thrombolytic; vulnerary; antitumor; osteopathic; immunosuppressive; antiinflammatory; cytostatic; gene therapy; chromosome 6.

Homo sapiens.

WO2003029271-A2.

10-APR-2003.

24-SEP-2002; 2002WO-US030474.

24-SEP-2001; 2001US-0324631P.

(HYSE-) HYSEQ INC.

Tang TY, Zhang J, Ren F, Xue AJ, Zhao QA, Wang J, Wehrman T; Zhou P, Ghosh M, Wang D, Ma Y, Asundi V, Wang Z, Weng G; Haley-Vicente D, Drmanac RT;

WPI; 2003-371981/35.

N-PSDB; ADC30642.

New polynucleotide and polypeptide useful for diagnosing, preventing or treating conditions such as neurodegenerative diseases, anemias, platelet disorders, wounds, ulcers, osteoporosis, autoimmune diseases or cancer.

Claim 20; SEQ ID NO 1695; 1185pp; English.

The invention relates to 971 novel human cDNA sequences (ADC29919-ADC30889) and the polypeptides they encode (ADC30890-ADC31860). The invention also relates to nucleic acid sequences over 99% identical with the novel human cDNAs. The invention additionally encompasses expression vectors and host cells comprising a nucleic acid of the invention; the recombinant production of a polypeptide of the invention; an antibody against a polypeptide of the invention; a method of detecting polynucleotides or polypeptides of the invention; and methods of identifying a compound which binds to a polypeptide of the invention. The invention further discloses methods of preventing, treating or ameliorating a medical condition; kits comprising polynucleotide probes and/or monoclonal antibodies for carrying out the methods of the invention; methods for the identification of compounds that modulate the expression or activity of the polynucleotide and/or polypeptide; and 767 contig sequences corresponding to the cDNA sequences of the invention (ADC31861-ADC32627) and the polypeptides encoded by the contigs (ADC32628-ADC33394). The nucleic acids and polypeptides of the invention are useful in diagnostics, drug screening, forensics, gene mapping, in the identification of mutations responsible for genetic disorders or other traits, for assessing biodiversity, and in producing many other types of data and products dependent on DNA and amino acid sequences. They are also used for treating diseases such as Parkinson's disease, Alzheimer's disease and other neurodegenerative diseases, anaemia, platelet disorders, wounds, ulcers, osteoporosis, autoimmune diseases or cancer. The nucleic acids may also be used as hybridisation probes or primers, and in the recombinant production of a protein. The polypeptides are also useful in generating antibodies, as molecular weight markers, and as food supplements. The present sequence represents a specifically claimed human polypeptide sequence of the invention. Note: The sequence data for this patent did not form part of the printed specification, but was obtained in electronic format directly from WIPO at [ftp.wipo.int/pub/published_pct_sequences](http://wipo.int/pub/published_pct_sequences).

Sequence 416 AA;

Query Match 24.4%; Score 659; DB 7; Length 416;
Best Local Similarity 44.2%; Pred. No. 5.3e-33;
Matches 136; Conservative 50; Mismatches 62; Indels 60; Gaps 2;

QY 281 KLRQNNMELAEERLKKLIEQVELREHIDKVPFKHDLQOOLVDAKLOQAQEMKKAERHQ 340
DB 2 KLCQENTELAEKLSIIDQVELREHIDKIFKRELOQKLVDKLEQAQEMKKAERHK 61
QY 341 REKDFLKAQVESQRMCELMKQOETHLKOOLALYTEKFEFQNTLSKSEVFTTFKQEME 400
DB 62 REKYLLNQAAEWKLOAKVLKSEIVLQAQUTLYSGRFEFQSTLTKSNEVFAFKQEMD 121
QY 401 KMTKKIKKLEKETTMYRSRWESSNKALLEMAEKTVDKLEGLQVQIKRLEKLCRALQT 460
DB 122 KTYTKMKKLEKDTATKARFENCNKALLDMIEEKALAAKEVECFVMKIGRLNLCRALQE 181
QY 461 ERNDLNKRVQDLSAGGQSLTDSQFERRPEG----- 491
DB 182 ERNELKKKIRDAEISEKDDQSQHSDEEPSNVSDQEIADAEVNSVQTAVNLTAFMI 241
QY 492 -----PCAQAPSSP-----RVTEAPCYPGAPSTASQOT 520
DB 242 IHPPESTPHQSKEQTEIGSSQESADAALKEPEQPLIPSRDSSPLPLTPQAEAGGS 301
QY 521 GPQEP TSA 528
DB 302 DAEPPSKA 309
RESULT 23
ABG02486
ID ABG02486 standard; protein; 410 AA.
AC ABG02486;
DT 13-FEB-2002 (first entry)
DE Novel human diagnostic protein #2477.
XX Human; chromosome mapping; gene mapping; gene therapy; forensic; food supplement; medical imaging; diagnostic; genetic disorder.
KW Homo sapiens.
OS WO200175067-A2.
PN 11-OCT-2001.
XX 30-MAR-2001; 2001WO-US008631.
PR 31-MAR-2000; 2000US-00540217.
PR 23-AUG-2000; 2000US-00649167.
XX (HYSE-) HYSEQ INC.
PI Drmanac RT, Liu C, Tang YT;
PI WPI; 2001-639362/73.
DR N-PSDB; AAS66673.
XX New isolated polynucleotide and encoded polypeptides, useful in diagnostics, forensics, gene mapping, identification of mutations responsible for genetic disorders or other traits and to assess biodiversity.
PT Claim 20; SEQ ID NO 32845; 103pp; English.
XX The invention relates to isolated polynucleotide (I) and polypeptide (II) sequences. (I) is useful as hybridisation probes, polymerase chain reaction (PCR) primers, oligomers, and for chromosome and gene mapping, and in recombinant production of (II). The polynucleotides are also used in diagnostics as expressed sequence tags for identifying expressed genes. (I) is useful in gene therapy techniques to restore normal activity of (II) or to treat disease states involving (II). (II) is useful for generating antibodies against it, detecting or quantitating a polypeptide in tissue, as molecular weight markers and as a food

CC supplement. (II) and its binding partners are useful in medical imaging
CC of sites expressing (II). (I) and (II) are useful for treating disorders
CC involving aberrant protein expression or biological activity. The
CC polypeptide and polynucleotide sequences have applications in
CC diagnostics, forensics, gene mapping, identification of mutations
CC responsible for genetic disorders or other traits to assess biodiversity
CC and to produce other types of data and products dependent on DNA and
CC amino acid sequences. ABB00010-ABG30377 represent novel human diagnostic
CC amino acid sequences of the invention. Note: The sequence data for this
CC patent did not appear in the printed specification, but was obtained in
CC electronic format directly from WIPO at
CC ftp.wipo.int/pub/published_pct_sequences

XX SQ Sequence 410 AA;
Query Match 23.8%; Score 642.5; DB 4; Length 410;
Best Local Similarity 40.9%; Pred. No. 5.6e-32;
Matches 153; Conservative 55; Mismatches 85; Indels 81; Gaps 11;
QY 167 STPEKLAALC-KKYABLL-EEHNSQKQKMLQKQS--QLVQKHLRGHSHKAV 219
DB 85 STMEE--AGLCGLREKADMLCSESH-----DILQHQDSNCSATSKHLLDEBGRDF 135
QY 220 LARSKLESLELQNRNLSKEGVQRAFEEREKKEVTSHFQVTLNDIQLQMEHNRN 279
DB 136 IYK-----NRSW-EHMQCKKEEVLKEVTAHFQITLTQALQEQHEIHN 181
QY 280 SKLRQENMELARLKLQYELREHIDKVPKHKDLOOQLVDKQLQQAQEMKKEAEERH 339
DB 182 AKLQENMEMGEKLLKLDQVALREQINKAFKHLRQLQVLDARLQQAQLINEADERH 241
QY 340 QREKDFLLKEAVESORMCMLKQOETHLQKQALVTERKFEFQNTLSKSSVEFTTQDEM 399
DB 242 QREREFLLKEAVESORMCMLKQOETHLQKQALVTERKFEFQNTLSKSSVEFTTQDEM 301
QY 400 EMTKKIKKLEKETTMYRSRWSSNNKALLEMAHEKTVTRDKLEGLQVKIQRLKLCRALQ 459
DB 302 EK-----XTIRDKYKVPQIKLERLEKLYKALQ 329
QY 460 TERNDLNKRV-----QDLSAGGQGLTDSGPER-----RPGPGAQAPSSR 501
DB 330 IERNELSEKGLLKGQSVKQVADVLAVPVTHSCADLSSNMKLTSSKRAFGVHEADPK 369
QY 502 -VTEAPCYGAPST 514
DB 390 GMEVCKYKALST 403

RESULT 24
ABB61894
ID ABB61894 standard; protein; 515 AA.
XX AC ABB61894;
XX DT 26-MAR-2002 (first entry)
XX DE Drosophila melanogaster polypeptide SEQ ID NO 12474.
XX KW Drosophila; developmental biology; cell signalling; insecticide;
XX KW pharmaceutical.
XX OS Drosophila melanogaster.
XX PN WO200171042-A2.
XX PD 27-SEP-2001.
XX PF 23-MAR-2001; 2001WO-US009231.
XX PR 23-MAR-2000; 2000US-0191637P.
XX PR 11-JUL-2000; 2000US-00614150.
XX PA (PEKE) PE CORP NY.

XX Venter JC, Adams M, Li PWD, Myers EW;
XX WPI; 2001-656860/75.
XX N-PSDB; ABL05997.
XX PT New isolated nucleic acid detection reagent for detecting 1000 or more
XX genes from Drosophila and for elucidating cell signalling and cell-cell
XX interactions.
XX PS Disclosure; SEQ ID NO 12474; 21pp + Sequence Listing; English.
XX CC The invention relates to an isolated nucleic acid detection reagent
XX capable of detecting 1000 or more genes from Drosophila. The invention is
XX useful in developmental biology and in elucidating cell signalling and
XX cell-cell interactions in higher eukaryotes for the development of
XX insecticides, therapeutics and pharmaceutical drugs. The invention
XX discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
XX sequences (ABU01840-ABU16175) and the encoded proteins (ABB57737-
XX ABB72072). The sequence data for this patent did not form part of the
XX printed specification, but was obtained in electronic format directly
XX from WIPO at ftp.wipo.int/pub/published_pct_sequences
XX SQ Sequence 515 AA;

Query Match 22.1%; Score 597.5; DB 4; Length 515;
Best Local Similarity 35.8%; Pred. No. 4.6e-29;
Matches 126; Conservative 98; Mismatches 105; Indels 23; Gaps 3;
QY 127 BEIROSDEVDGDRHRRPQKKAKGLGKEITLMTQTLNLTSPPEKLAALCKYAEILLE 186
DB 2 EKLSKAKVAREKORDKLE-----LVMSLDECPSAEKVKL-----LLQR 45
QY 187 HNSOK-----QMKLQKQSQVLQVKHDLRGHSHKAVLARSKLESLELQNRNLS 239
DB 46 HYDSEKNVSRLLFAELRVLRQMESQOREKEQVQRLNKLNRDLQVCEQOQRIKSV 105
QY 240 KEEGVQRAFEEREKKEVTSHFQVTLNDIQLQMEHNRNLSKLRQENMELARLKLIEQ 299
DB 106 KNESSLQIKVSEERKESQTKFQSSLDNDVQKSLAGNNEIKLDYNIEMTKLKLAAEQ 165
QY 300 YELREHIDKVPKHKDLOOQLVDKQLQQAQEMKKEAEERHOREKDFLKEAVESORMCEL 359
DB 166 YQTRQHLKLEKNEQVLEAQLHQAQKQKQVFAAMEKEILSKENQIGLEKLMQARAID 225
QY 360 MKQOETHLQKQALVTERKFEFQNTLSKSSVEFTTQDEMKTCKIKKLEKETTMYRSR 419
DB 226 LTDRHQLEKQLNIYTKYDDFQQSLOKSNVFGSYKVELKSKSKYTKIEKALGHRQK 285
QY 420 WESSNKALLEMAHEKTVTRDKLEGLQVKIQRLKLCRALQTERNDLNKRVQD 471
DB 286 YEKANAMVIDLATERKSLQTHSERLQKIQOQLKLLRALQLERTTLHKCLRD 337

RESULT 25
ABG02487
ID ABG02487 standard; protein; 841 AA.
XX AC ABG02487;
XX DT 13-FEB-2002 (first entry)
XX DE Novel human diagnostic protein #2478.
XX KW Human; chromosome mapping; gene mapping; gene therapy; forensic;
XX KW food supplement; medical imaging; diagnostic; genetic disorder.
XX OS Homo sapiens.
XX PN WO200175067-A2.
XX PD 11-OCT-2001.
XX

30-MAR-2001; 2001WO-US008631.	
31-MAR-2000; 2000US-00540217.	
23-AUG-2000; 2000US-00649167.	
(HYSE-) HYSEQ INC.	
Dmanac RT, Liu C, Tang YT;	
WPI; 2001-639362/73.	
N-PSDB; AAS66674.	
New isolated polynucleotide and encoded polypeptides, useful in diagnostics, forensics, gene mapping, identification of mutations responsible for genetic disorders or other traits and to assess biodiversity.	
Claim 20; SEQ ID NO 32846; 103pp; English.	
The invention relates to isolated polynucleotide (I) and polypeptide (II) sequences. (I) is useful as hybridisation probes, polymerase chain reaction (PCR) primers, oligomers, and for chromosome and gene mapping, and in recombinant production of (II). The polynucleotides are also used in diagnostics as expressed sequence tags for identifying expressed genes. (I) is useful in gene therapy techniques to restore normal activity of (II) or to treat disease states involving (II). (II) is useful for generating antibodies against it, detecting or quantitating a polypeptide in tissue, as molecular weight markers and as a food supplement. (II) and its binding partners are useful in medical imaging of sites expressing (II). (I) and (II) are useful for treating disorders involving aberrant protein expression or biological activities. The polypeptide and polynucleotide sequences have applications in diagnostics, forensics, gene mapping, identification of mutations responsible for genetic disorders or other traits to assess biodiversity and to produce other types of data and products dependent on DNA and amino acid sequences. ABG00010-ABG30377 represent novel human diagnostic amino acid sequences of the invention. Note: The sequence data for this patent did not appear in the printed specification, but was obtained in electronic format directly from WIPO at ftp.wipo.int/pub/published_pct_sequences	
Sequence 841 AA;	
Query Match	20.4%; Score 550; DB 4; Length 841;
Best Local Similarity	50.0%; Pred. No. 7.4e-26;
Matches 121; Conservative 40; Mismatches 51; Indels 30; Gaps 7;	
167 STPEKLAALC--KKYAEIL---EHNRSQMKMLQKKQS--QLVQEKDHLRGEHSAV 219	: : : : : : : : : : : : : :
53 STMGE--AGLCGLRKADMLCNSESH-----DILQHQDNCATSATSKHLLDEGEGRDF 103	: : : : : : : : : : : : :
220 LARSKLESICRELQRNHRSLKEGVQVAREEERKKEVTSHFQVTLNDIQLQWEGHNERN 279	: : : : : : : : : : : : :
104 ITK-----NRSW-EHHMQQKKEEEVLKEVTAHQITLTETQAQLQEQHEITHN 149	: : : : : : : : : : : : :
280 SKLQENKELAERLKLLIEQVELREEHIDKVFVGHKLQQQLVDAKLQQAQEMLKAEERH 339	: : : : : : : : : : : : :
150 AKLAQENNMMEGKLKCLTDQYALREQINKAFPHKELRQQLVDARLQQTAQIKAEERH 209	: : : : : : : : : : : : :
340 QREKDFLLKAEVRSORMCELMKQOETHLKQALALYTEKPEEPONTLSKSEVFTTFKOBM 399	: : : : : : : : : : : : :
210 QREREFLLKATEPSRHKYEEKQAEQKQLFLYMDKPEEFQTTWAKTNELFTAFKQET 269	: : : : : : : : : : : : :
400 EK 401	
270 EK 271	
RESULT 26	
ABG02488	
IID ABG02488 standard; protein; 102 AA.	
XX	
XX ABG02488;	
XX	

Db 565 KVAVSPTPPVPRSTFVAPPEQSLSEALKAMBEAQAQVLE-----QDQHLLSKOE 619
QY 203 QLVQEKDHLRGEHSKAVLA--RSKLSLCLREIQHNRSLSKEGVQARBEK-----RK 255
Db 620 KMQLREKLQCEBEBILHQQKEQSLSLRERLQKATEEBA--RMRBEESQRLSWLRA 678
QY 256 EVTSHFQVTLNDLQONE-----QHNERNKLQENMELAEKLIQOYELR 303
Db 679 QVQSSSTQADEQIRACQERASLOKLRELSQQAERASLEQKRMQLKKEIEASEKS 738
QY 304 BEHIDKVPFKDLQO---QLVDAKLOQAQEMLE-----ABERHOREKDFLLKE 349
Db 739 EQAALNAAKKALQQLREQLGEBRKEAVATLEKHSABLERLCSLEAKHREVVSLQKK 798
QY 350 AVESQRMCLMKQ---QETHLQQLALYTEKPEEFQNTL--SKSEVETTFKQEMKMT 403
Db 799 IOAQOQKEAQKQKCLGQVEHRVHQSYHVAGYEHLSLLREKREQVEGHEHRLDRM- 857
QY 404 KTKIKLEKETTMYRSWESSNKALLEWABEKTVDKXELGLOVKIQRLKLC--RALQTER 462
Db 858 ----KEHQQVMKAREQYE-----ABERKQASLLGHLTQELERLQRAHRELETVR 906
QY 463 NDLNKRVDL 472
Db 907 QEQHKELEDL 916

RESULT 30
ADC38517
ID ADC38517 standard; protein; 860 AA.
XX
AC ADC38517;
XX
DT 18-DEC-2003 (first entry)
XX
DE Human AMLP1 myosin-tail motif consensus protein SEQ ID NO:866.
XX
KW human; angiominotin-like protein 1; AMLP1; cytostatic; gene therapy.
XX
OS Synthetic.
OS Homo sapiens.
XX
XX WO2003037931-A2.
XX
PD 08-MAY-2003.
XX
PF 01-NOV-2002; 2002WO-US035129.
XX
PR 01-NOV-2001; 2001US-0334773P.
XX
PA (AMSH) AMERSHAM BIOSCIENCES SV CORP.
XX
PI Shannon M, Phan T;
XX
XX WPI; 2003-430501/40.
XX
XX New isolated nucleic acid molecule encoding a human angiominotin-like
PT protein, useful for treating or preventing a disorder associated with
PT decreased or increased expression or activity of AMLP1.
XX
XX Example 2; SEQ ID NO 866; 172pp; English.
XX
XX The present invention describes the human angiominotin-like protein 1
CC (AMLPI). human AMLP1 has cytostatic activity and can be used in gene
CC therapy. The AMLP1 protein, nucleic acid molecules, antibodies, and
CC compositions of the present invention can be used for treating or
CC preventing a disorder associated with decreased or increased expression
CC or activity of AMLP1. The present sequence represents a consensus myosin-
CC tail motif amino acid sequence, which is used in an example from the
CC present invention.
XX
XX Sequence 860 AA:
SQ

Query Match 9.5%; Score 257.5; DB 7; Length 860;
Best Local Similarity 22.7%; Pred. No. 1.2e-07;
Matches 143; Conservative 106; Mismatches 189; Indels 193; Gaps 25;
QY 12 EGAEPRQQAAPAYEAEGFGSSQAPRKEPGEQAQTAQSGALRDVSELSRQLEDILSYTC 71
Db 77 EELSERLEBEGGATAAQ-----LELNKKEAEALAKLRKDLERANLQHEFALAT-L 125
QY 72 VDNQGGPGEQGAPEPAEPEDAESKRTYVARNGEPPEPTVYGEKPSKGDPTNEB-IR 130
Db 126 RKHQDAINELSEIQLOKQAK-----LLMOTLNTLSTPEEKJAA- 166
QY 131 QSDEVDGRDHRPPOEKKKAGLGKEIT-----LLMOTLNTLSTPEEKJAA- 175
Db 167 QLDSI---TKAIAAEKAKQLESQLSQVLEKLBQRLNDLTSKRLQSENSDLTRQ 223
QY 176 -----LCKYAEI---LEHRNS-----QKMKLLQKQSQVQEKDHLRGEHKA 218
Db 224 LEBAQVSNLSKLQSLESQLEAEKGLSEESERANLQALQLEHDLQSLRQLEEE 283
QY 219 VLARSKLESQLELQHNRSIKE-----EGVQARPEERKEKVT----- 258
Db 284 SEAKAELE---RLSKANAEIQWRSKFESESGALRAEELBLKKLNQKISELEBAEAA 340
QY 259 -----SHFOVTINDIQLQWQHNNRSKI---ROENMB--LABRLKKLI-E----- 298
Db 341 NAKCDSLEKTKSLQSELEDLQIELERANAAASBLEKKQKQKFDKILAEWKVKVDELQAE 400
QY 299 ---QVELR-----BEHIDKV---FKHKLQOQQLVDKLAQQAQ-----EMLK 333
Db 401 DTAGREARNLSTELFKKNELEELKQVEALRRKNLQDSIHDLTDQLGEGGRNVHELE 460
QY 334 BAEERHOREKDFLLKEAVESQRMCEL-----MKQETHLQKQALALYTEKPEFPQ 382
Db 461 KARRLEAEKDELQAALEBAEAALEBESKVLRAQVLSQIRSEIERLA---EKBEFE 517
QY 383 NT-----LSKSSEVETTFKQEMKMTKK-----IK 407
Db 518 MTRKNHQRAIESLOATLEATGKAEASRLKKLEGDINELEIALDHANKANAEAKQNVK 577
QY 408 KLEKETTMYRSWRSSNKA-----LLEMAEEK-TVRDKLEGLQVKTQRLKLCRALQOT 460
Db 578 KYQQQVKELQTVBEEQARADAREQLAVAEERRATALEAELESELSALEQAERARKQAE 637
QY 461 ERNDLNKQVQLSAGCGQSLTDSGPEREPG 491
Db 638 ELAEASERVNELTAQNSSLIAQ---KKKLEG 665

RESULT 31
AAY07031
ID AAY07031 standard; protein; 1752 AA.
XX
AC AAY07031;
XX
DT 02-JUL-1999 (first entry)
XX
DE Breast cancer associated antigen precursor sequence.
XX
KW Cancer associated antigen; diagnosis; research; treatment; human;
KW breast cancer; colon cancer; gastric cancer; renal cancer; lung cancer;
KW prostate cancer.
XX
OS Homo sapiens.
XX
PN WO9904265-A2.
XX
PD 28-JAN-1999.
XX
PF 15-JUL-1998; 98WO-US014679.
XX
PR 17-JUL-1997; 97US-00896164.
PR 10-OCT-1997; 97US-0061599P.

PR 10-OCT-1997; 97US-0061765P.
PR 10-OCT-1997; 97US-00948705.
PR 11-OCT-1997; 97GB-00021697.
PR 22-JUN-1998; 98US-00102222.
XX (LUDW-) LUDWIG INST CANCER RES.
XX
PI Old LJ, Scanlan WJ, Stockert E, Gure A, Chen Y, Gout I;
PI O'hare M, Obata Y, Pfeundschnuh M, Tureci O, Sahin U;
XX WPI; 1999-132448/11.
XX
PT New isolated cancer associated nucleic acids and polypeptides - isolated
PT using sera from cancer patients, used to develop products for the
PT diagnosis, monitoring or treatment of cancers.
XX
XX Disclosure; Page 409-413; 787pp; English.
XX
CC The invention relates to a method for diagnosing a disorder characterised
CC by expression of a human cancer associated antigen precursor coded for by
CC a nucleic acid molecule (NAM). The method comprises: (a) contacting a
CC biological sample isolated from a subject with an agent that specifically
CC binds to the NAM, an expression product or a fragment of an expression
CC product complexed with an HLA molecule; and (b) determining the
CC interaction between the agent and the NAM or the expression product as a
CC determination of the disorder. The products and methods can be used in
CC the diagnosis, monitoring, research, or treatment of conditions
CC characterised by the expression of various cancer associated antigens.
CC The invention provides nucleic acid sequences and encoded polypeptides
CC which are cancer associated antigen precursors expressed in human breast
CC cancer, renal cancer, colon cancer, gastric cancer, prostate cancer and
CC lung cancer
XX
XX Sequence 1752 AA;
Query Match 9.5%; Score 256; DB 2; Length 1752;
Best Local Similarity 23.2%; Pred. No. 3.5e-07;
Matches 133; Conservative 107; Mismatches 195; Indels 138; Gaps 25;
QY 26 EAGPGSSQAPRP-----EGNARTAGS-----GALADVSELSQEDIL 67
DB 313 QKATQSEVSRKQQLLEVLRTQVMTBESVRYKOSLDAAKTIQDNKBIER-LQLLI 371
QY 68 STYCVNNNGGPDGAGQAPFAPEDAEKSRIT-YVARNGEPPEPTPVV---YGEKPSKG 122
DB 372 DKEINR-----KLEDENARLQVQYDLQKANSATETINKLKVQEBELTRL 419
QY 123 DPNTETIRQDEVDHRPQPKKAKGLGKELITLLMOTLNTSTPEEKLAALCKKYAB 182
DB 420 RIDYERSQERTVKDQDITRFQNSLRELQLOKQ--KVEEELNKLKRTASDSCCKKLES 477
QY 183 LLEHRNSQOMKL-----LOKQSQ--LVQSKD---HLR-GEHSAVLAR 222
DB 478 ELEGWRSLEQAIKITNLTOQLEQASIVKRSDDLQQRDVLGHLREKQRTQBELR 537
QY 223 --SKLSLCLRELQHNHSLKE-----EGVQARABEBEKKKVTSHFQVTLNDIQLQMHQH 275
DB 538 LSSEVEALRQLLQEQESVQAHLNEHFQKATEDKSR-----SLNESKIBIERL 587
QY 276 NERNKSLQENMELARLKL-TEQVELR-----BEHIDK-----VPKHKDLQ 318
DB 588 QSLUTNUTKHELMEEELNRLREYDRLRGREASDKNATILELSQIISNNITL 647
QY 319 QLVDAKLQQAQEMKEAEERHQEKDFLLKEAVESQRM-C-ELMKQETHLQQLALYTEK 377
DB 648 QGLINDLQREMLRQBIETKQALASNRIOESKNOCTQVQVERESLLVKKVLEQDK 707
QY 378 -----FEFQNTLKSSEVETTKQEMKMKKKK-LEKETTYMRSSNKKALLEMA 431
DB 708 ARQLRLEDEINAKSTLEATRVKORLECEKQIONDNQWTKYGRKEBAIRK--IESE 765
QY 432 EEKTVRDK-----ELEGLQVKIORLEKLCR-----ALQTERNDLNKRVQDLGA 474

DB 766 REKSRKNSLASEIERLQAEIKRIEBCRKLESTRETQSLTERSRVORIDKL-- 823
QY 475 GGQGSILTDGPPRRPPEGCAQAPSSPRVTEAPC 507
DB 824 -----RQRPYG-----SHRETQTEC 838
RESULT 32
ABU07402
ID ABU07402 standard; protein; 2871 AA.
XX AC ABU07402;
XX DT 28-JAN-2003 (first entry)
XX DE Protein differentially regulated in prostate cancer #5.
XX KW Prostate cancer; gene expression; differential regulation;
KW molecular marker; drug target; cancer detection; cancer diagnosis;
KW cancer staging; cancer grading; cancer assessing; cancer monitoring.
XX OS Homo sapiens.
XX PN WO200281638-A2.
XX PD 17-OCT-2002.
XX PF 08-APR-2002; 2002WO-US010824.
XX PR 06-APR-2001; 2001US-0281731P.
XX PR 06-APR-2001; 2001US-0281732P.
XX PA (ORIG-) ORIGENE TECHNOLOGIES INC.
XX PI Sun Z, Jay G;
XX DR WPI; 2003-058520/05.
XX PT Novel genes which are differentially regulated in prostate cancer, useful
XX for diagnosing prostate cancer in prostate tissue sample and assessing
XX therapeutic or preventive intervention in prostate cancer patients.
PS Claim 1; Page 209-219; 416pp; English.

The invention describes genes (I) which are differentially regulated in prostate cancer. (I) is useful for diagnosing a prostate cancer in a sample comprising prostate tissue, which involves determining the number of target genes which are differentially-regulated in the sample, where the number is indicative of the probability that the sample comprises prostate cancer. (I) is useful for assessing a therapeutic or preventive intervention in a subject having a prostate cancer, which involves determining the expression levels in a sample comprising prostate tissue of target genes which are differentially-regulated in prostate cancer. Preferably, the expression levels of at least 10 genes are determined. (I) is also useful for identifying agents that modulate a biological activity of a polypeptide differentially-regulated in prostate cancer cells, which involves contacting a polypeptide differentially-regulated in prostate cancer cells with a test agent under conditions effective for the test agent to modulate a biological activity of the polypeptide, and determining whether the test agent modulates the biological activity. (I) is useful as molecular markers, as drug targets, and for detecting, diagnosing, staging, grading, assessing, monitoring, prognosticating, preventing or treating, determining predisposition to diseases and conditions especially relating to prostate cancer. (I) and its expression products are used in the diagnostic test to assay for presence of cancer e.g., in tissue sections, in biopsy sample, in total RNA, in lymph, in blood etc. (I) is useful for assessing cancer e.g., to determine the type of cancer, its stage of development, the nature of genetic defect, etc. The polypeptide encoded by (I) can be used as target for therapy or drug discovery. (I) can also be used for expressing the polypeptide and thus for searching specific binding partners of the polypeptide. (I) is useful in therapeutic applications to treat prostate cancer. The identification of specific genes, and groups of genes, expressed in pathways

CC physiologically relevant to prostate cancer permits the definition of
CC functional and disease pathways and the delineation of targets in these
CC pathways which are useful in diagnostic, therapeutic, and clinical
CC applications. This is the amino acid sequence of a protein differentially
CC regulated in prostate cancer
XX
XX
SQ Sequence 2871 AA;

Query Match 9.5%; Score 256; DB 6; Length 2871;
Best Local Similarity 23.2%; Pred. No. 6.2e-07;
Matches 133; Conservative 107; Mismatches 195; Indels 138; Gaps 25;
QY 26 BAEGPSSQAPRKP-----EGQAARTAQ-----GALRDVSELSQLDIL 67
DB 1432 QKATGSEVSQKQQLVLELQVMTREESVRYKSLDDAAKTIQDKNEIER-LKQLI 1490
QY 68 STYVDNNGGPGEDGAGPAPEDAEST-YVANGPEPTPVV-----YGEKPSKG 122
DB 1491 DKETNDR-----KCLEENARLQVQVLDQKANSATETINKLVQEQELTRL 1538
QY 123 DPNTBEIRQSDVGDHRRPQKKKAGLGEITLLMTLSTPEEKLAALCKKYAE 182
DB 1539 RIDYERVSQERTVKDDITRFQNSLKLQLOKQ--KVVEELNRLKRTASEDSCKRKL 1596
QY 183 LLEHRNSQKMKL-----LQKQSQ--LVQEKD-----HLR-GEHSAVLAR 222
DB 1597 ELEGRRSLKEQAIIKITNLTOLEQASIVKRSDDLQQRDVLGDHLEKQRTQELRR 1656
QY 223 --SKLESCLRELQHNLSLKE-----EGVQARBEERKEKVTSHFQVTLNDILOMEQH 275
DB 1657 LSSEVALRRQLQEQESVKQAHLEHFKAIEDKSR-----SLNESKIEIERL 1706
QY 276 NERNKSLRQENMELAEKLL-IEQVELR-----EHIK-----VFKHDLQ 318
DB 1707 QSLTENITKEHLMLEELNRLLEYDRLRRGSEADSKNATILELSQLQISNNRTLEL 1766
QY 319 QLVDAKLAQQAQEMLEKEAERHREKDFLLKEAVESQMC-ELMKQOETHLQQLALYTEK 377
DB 1767 QGLINDLQRENLRLQBIKFKQALEASNRIOESKNQCTQVQVERSLLVKIKVLEQDK 1826
QY 378 -----FEFPQNTLSKSSEVFTTFKQEMKMTKIKK-LEKETTYRWSWSSNKALLEMA 431
DB 1827 ARQLRDELNRAKSTLEAEITRVKQRLCEKQIQNDLQNMKTQYSRKEBAIRK--IESE 1884
QY 432 EKTVRDK-----ELEGQVKIQRLKLCR-----ALQTERNDLNKRVQDLSA 474
DB 1885 REKSERKNSLRSEIERLQAEIKRIEERCCKLEKLDSTRETOSQLETYSRYQREIDKL-- 1942
QY 475 GGQSLTDSGPERPEGPGCAQAPSPRVTEAPC 507
DB 1943 -----RQPYG-----SHRETQTEC 1957

RESULT 33

ADC35075
ID ADC35075 standard; protein; 2871 AA.
XX
XX ADC35075;
XX
XX
DT 18-DEC-2003 (first entry)
XX
XX Human breast cancer antigen seq id 41.
XX
XX breast cancer; breast cancer diagnosis; breast cancer antigen.
XX
XX Homo sapiens.
XX
XX US2003108888-A1.
XX
XX 12-JUN-2003.
XX
XX 15-MAY-2002; 2002US-00146473.

PR 15-MAY-2001; 2001US-0291150P.
XX
XX (LUDW-) LUDWIG INST CANCER RES.
XX
XX Scanlan MJ, Gout I, Stockert B, Old LJ, Gure A, Chen Y;
XX
XX WPI; 2003-829397/77.
XX
XX N-PSDB; ADC35117.
XX
XX Diagnosing breast cancer in subject by obtaining biological sample from
XX
XX subject, contacting sample with breast cancer-associated polypeptides,
XX
XX determining specific binding between polypeptides and agents in sample.
XX
XX Example 2; SEQ ID NO 41; 173pp; English.
XX
XX The invention describes a method of diagnosing breast cancer in subject
XX
XX comprising contacting biological sample from subject with at least two
XX
XX different breast cancer-associated polypeptides (I) encoded by nucleic
XX
XX acid molecules (II) comprising sequence chosen from 42 fully defined
XX
XX sequences as given in specification, determining specific binding between
XX
XX (I) and agents in sample, where presence of the binding is diagnostic for
XX
XX breast cancer. The method is useful for diagnosing breast cancer in a
XX
XX subject. The sample is blood, lymph node fluid or breast discharge fluid.
XX
XX This is the amino acid sequence of a breast cancer antigen.
XX
XX Sequence 2871 AA;

Query Match 9.5%; Score 256; DB 7; Length 2871;
Best Local Similarity 23.2%; Pred. No. 6.2e-07;
Matches 133; Conservative 107; Mismatches 195; Indels 138; Gaps 25;
QY 26 BAEGPSSQAPRKP-----EGQAARTAQ-----GALRDVSELSQLDIL 67
DB 1432 QKATGSEVSQKQQLVLELQVMTREESVRYKSLDDAAKTIQDKNEIER-LKQLI 1490
QY 68 STYVDNNGGPGEDGAGPAPEDAEST-YVANGPEPTPVV-----YGEKPSKG 122
DB 1491 DKETNDR-----KCLEENARLQVQVLDQKANSATETINKLVQEQELTRL 1538
QY 123 DPNTBEIRQSDVGDHRRPQKKKAGLGEITLLMTLSTPEEKLAALCKKYAE 182
DB 1539 RIDYERVSQERTVKDDITRFQNSLKLQLOKQ--KVVEELNRLKRTASEDSCKRKL 1596
QY 183 LLEHRNSQKMKL-----LQKQSQ--LVQEKD-----HLR-GEHSAVLAR 222
DB 1597 ELEGRRSLKEQAIIKITNLTOLEQASIVKRSDDLQQRDVLGDHLEKQRTQELRR 1656
QY 223 --SKLESCLRELQHNLSLKE-----EGVQARBEERKEKVTSHFQVTLNDILOMEQH 275
DB 1657 LSSEVALRRQLQEQESVKQAHLEHFKAIEDKSR-----SLNESKIEIERL 1706
QY 276 NERNKSLRQENMELAEKLL-IEQVELR-----EHIK-----VFKHDLQ 318
DB 1707 QSLTENITKEHLMLEELNRLLEYDRLRRGSEADSKNATILELSQLQISNNRTLEL 1766
QY 319 QLVDAKLAQQAQEMLEKEAERHREKDFLLKEAVESQMC-ELMKQOETHLQQLALYTEK 377
DB 1767 QGLINDLQRENLRLQBIKFKQALEASNRIOESKNQCTQVQVERSLLVKIKVLEQDK 1826
QY 378 -----FEFPQNTLSKSSEVFTTFKQEMKMTKIKK-LEKETTYRWSWSSNKALLEMA 431
DB 1827 ARQLRDELNRAKSTLEAEITRVKQRLCEKQIQNDLQNMKTQYSRKEBAIRK--IESE 1884
QY 432 EKTVRDK-----ELEGQVKIQRLKLCR-----ALQTERNDLNKRVQDLSA 474
DB 1885 REKSERKNSLRSEIERLQAEIKRIEERCCKLEKLDSTRETOSQLETYSRYQREIDKL-- 1942
QY 475 GGQSLTDSGPERPEGPGCAQAPSPRVTEAPC 507
DB 1943 -----RQPYG-----SHRETQTEC 1957

RESULT 34

ABG27218
ID ABG27218 standard; protein; 29:8 AA.
XX
AC ABG27218;
XX
DT 18-FEB-2002 (first entry)
XX
DE Novel human diagnostic protein #27209.
XX
KW Human; chromosome mapping; gene mapping; gene therapy; forensic;
KW food supplement; medical imaging; diagnostic; genetic disorder.
XX
OS Homo sapiens.
XX
FN WO200175067-A2.
XX
PD 11-OCT-2001.
XX
PF 30-MAR-2001; 2001WO-US008631.
XX
PR 31-MAR-2000; 2000US-00540217.
PR 23-AUG-2000; 2000US-00649167.
XX
PA (HYSE-) HYSEQ INC.
XX
PI Drmanac RT, Liu C, Tang YT;
XX
DR WPI; 2001-639362/73.
DR N-PSDB; AAS91405.
XX
XX New isolated polynucleotide and encoded polypeptides, useful in
PT diagnostics, forensics, gene mapping, identification of mutations
PT responsible for genetic disorders or other traits and to assess
PT biodiversity.
XX
PS Claim 20; SEQ ID NO 57577; 103pp; English.
XX
CC The invention relates to isolated polynucleotide (I) and polypeptide (II)
CC sequences. (I) is useful as hybridisation probes, polymerase chain
CC reaction (PCR) primers, oligomers, and for chromosome and gene mapping,
CC and in recombinant production of (II). The polynucleotides are also used
CC in diagnostics as expressed sequence tags for identifying expressed
CC genes. (I) is useful in gene therapy techniques to restore normal
CC activity of (II) or to treat disease states involving (II). (II) is
CC useful for generating antibodies against it, detecting or quantitating a
CC polypeptide in tissue, as molecular weight markers and as a food
CC supplement. (II) and its binding partners are useful in medical imaging
CC of sites expressing (II). (I) and (II) are useful for treating disorders
CC involving aberrant protein expression or biological activity. The
CC polypeptide and polynucleotide sequences have applications in
CC diagnostics, forensics, gene mapping, identification of mutations
CC responsible for genetic disorders or other traits to assess biodiversity
CC and to produce other types of data and products dependent on DNA and
CC amino acid sequences. ASG00010-ASG30377 represent novel human diagnostic
CC amino acid sequences of the invention. Note: The sequence data for this
CC patent did not appear in the printed specification, but was obtained in
CC electronic format directly from WIPO at
CC ftp.wipo.int/pub/published_pct_sequences
XX
SQ Sequence 2918 AA;
XX
Query Match 9.5%; Score 256; DB 4; Length 2918;
Best Local Similarity 23.2%; Pred. No. 6.3e-07;
Matches 133; Conservative 107; Mismatches 195; Indels 138; Gaps 25;
XX
QY 26 EAGGPGSQAPRKP-----EQAQTAQS-----GALRDVSELSROLEDIL 67
XX
DB 1479 QKATGSEVSQRKQQLVELRQVQMTTEESVRYKQSLDDAAKTQDNKRIER-LKOLI 1537
XX
QY 68 STYCVNNQGGPGEGCAGCEPAEPDAAKSRF-YVARNGBEPPTPV-----YGRKEPSKG 122
XX
DB 1538 DKETNDR-----KCLEENARLQRYQYDOLQKANSATETINKLVQEOELTRL 1585

QY 123 DPNTTEIRQSDVGDGRDRRPPQKKQAGKGLGKBITLMTLNTLSTPEKLAALCKYAE 182
DB 1586 RIDYRVUSQERTVKQDITRFQNSLKEQLQKQ--KVEELNRLKRTASEDSCKRKLKE 1643
QY 183 LLEHRNSQKQKML-----LQKQSQ--LVQEKD-----HLR-GEHSKAVLAR 222
DB 1644 ELEGMRSLKEQAIKITNLTOQLSQASIVKGRSDDLRRQQRDVLGHLRQKRTQOELRR 1703
QY 223 --SKLESICRELQRHNRSLKE-----EGVQPARREBEKKEVTSHFQVTLNDIQLQMPQH 275
DB 1704 LSSEVEALRQLLOQBSVKQALHRLNEHFQAIKDKSR-----SLNESKIEIRL 1753
QY 276 NERNSKLROENMELARLKKL-IEQYELR-----BEHIDK-----VFKHOLQQ 318
DB 1754 QSLTENLTKEHLMLEELRLNRLLEYDLRGRSEADSKKATILLELSQLQISNNRTLEL 1813
QY 319 QLVDAKLQAOQEMLKABERHOREKDFLLKSAVESQRMW-ELMKQOQTHLQOQLALVTEK 377
DB 1814 QGLINDLQREKENLRQIEKFKQALASNRIBQSKNOCTQVQVERSLVIVKIVLQDK 1873
QY 378 -----FEFQNTLSKSSEVFTTFKQEMKMTKIKK-LEKETTMYRSRWSSNKALLEMA 431
DB 1874 ARLQRLDELNRAKSTLEAETRVKQRLCEKQKQIQNDLNQWKTQYSRKEBAIRK--IESE 1931
QY 432 EEKTVRK-----ELEGLOVKIQLEKLCR-----ALQTERNDLNKQVQLSA 474
DB 1932 REKSERKNSLRSEIERLQAEIKKIERCRKLEKLESTRETQSQLETERSRYQREIDKL-- 1989
QY 475 GGQSLTDSQFERRPFGGAQAPSSPRVTEAPC 507
DB 1990 -----RQRPYG-----SHRETQTEC 2004
XX
RESULT 35
AAM79504
ID AAM79504 standard; protein; 931 AA.
XX
AC AAM79504;
XX
DT 06-NOV-2001 (first entry)
XX
DE Human protein SEQ ID NO 3150.
XX
KW Human; cytokine; cell proliferation; cell differentiation; gene therapy;
KW vaccine; peptide therapy; stem cell growth factor; haematopoiesis;
KW tissue growth factor; immunomodulatory; cancer; leukaemia;
KW nervous system disorder; arthritis; inflammation.
XX
OS Homo sapiens.
XX
FN WO200157190-A2.
XX
PD 09-AUG-2001.
XX
PF 05-FEB-2001; 2001WO-US004098.
XX
PR 03-FEB-2000; 2000US-00496914.
PR 27-APR-2000; 2000US-00560875.
PR 20-JUN-2000; 2000US-00598075.
PR 19-JUL-2000; 2000US-00620325.
PR 01-SEP-2000; 2000US-00654936.
PR 15-SEP-2000; 2000US-00663561.
PR 20-OCT-2000; 2000US-00693325.
PR 30-NOV-2000; 2000US-00728422.
XX
PA (HYSE-) HYSEQ INC.
XX
PI Tang YT, Liu C, Drmanac RT, Asundi V, Zhou P, Xu C, Cao Y;
PI Ma Y, Zhao QA, Wang D, Wang J, Zhang J, Ren F, Chen R, Wang ZH;
PI Xue AJ, Yang Y, Wejhrman T, Goodrich R;
XX
DR WPI; 2001-476283/51.
DR N-PSDB; AAK52637.

Db 643 EKTEALEQSELERENRKLKTLDS-----FKNLTFQLESLEKENSQLOEENLERNNV 697
QY 294 KXIEQVELFEEDHIDKFKHKLQOOLVDAKLQAOEMLKEAEERHOREKDFLLKEA--- 350
Db 698 ESL-----KCSMKMQLQLENKELSESEKQKGLG-----LLKASPKK 737
QY 351 -----VESQRMCLMKQOQETHLQOALYTEKPEEPQNTLSKSESVFTTFKQE 398
Db 738 TERLEVSQGLDIENQRLQTLNSNKKI-QQLESELQDLEMNQTLQKNLELKLSSKR 796
QY 399 MEKWTKKIKKLEKETTMY---RSRWESSNALLEMAEETKVRDKELEGQVQIQRKLKC 455
Db 797 LEQLEKENKSLBQETSQLEKDKQKQLEKENRKLQQAEB---IKDTTLENNVKLGNLEKEN 853
QY 456 RALQTE 461
Db 854 KTLKSE 859

RESULT 37
ABP73809
ID ABP73809 standard; protein; 1881 AA.
XX AC
XX ABP73809;
XX 30-JAN-2003 (first entry)
XX Candida albicans essential protein SEQ ID NO 7646.
XX Fungus; yeast; tetracyclin; promoter; GRACE strain; biosynthesis;
XX signal transduction; DNA replication; cell division; growth;
XX proliferation; Candida albicans; fungicide; antifungal.
XX Candida albicans.
XX OS
XX W0200253728-A2.
XX 11-JUL-2002.
XX 26-DEC-2001; 2001WO-US049486.
XX 29-DEC-2000; 2000US-0259128P.
XX 20-FEB-2001; 2001US-00792024.
XX 22-AUG-2001; 2001US-0314050P.
XX (ELIT-) ELITRA PHARM INC.
XX Roemer T, Jiang B, Boone C, Bussey H, Ohlsen KL;
XX WPI; 2002-566694/60.
XX DR N-PSDB; ABZ32359.
XX Constructing strains for identifying gene products as effective targets
XX for therapeutic intervention, by inactivating in the strain one allele of
XX a gene and placing other allele of the gene under conditional expression.
XX Claim 44; SEQ ID NO 7646; 167pp + Sequence Listing; English.
XX The invention relates to constructing (M1) a strain of diploid fungal
XX cells in which both alleles of a gene are modified, comprising modifying
XX one allele by insertion or replacement by a cassette having an
XX expressible selectable marker and modifying other allele by
XX recombination, of a promoter replacement fragment with a heterologous
XX promoter, so that expression of the second allele is regulated by the
XX promoter. (M1) is useful for constructing a strain of diploid fungal
XX cells in which both alleles of a gene are modified. The diploid fungal
XX cells having both alleles modified are useful for identifying a gene that
XX is essential to the survival or growth of a fungus, a gene that
XX contributes to the virulence and/or pathogenicity of a fungus, a gene
XX that contributes to the resistance of a diploid fungus to an antifungal
XX agent, an antifungal agent that inhibits the growth of a diploid fungus
XX and for identifying a therapeutic agent for treatment of a mammalian

CC disease. (M1) is useful for identifying a compound which modulates the
CC activity of a gene product, preferably enzymatic activity, carbon
CC compound catabolism, biosynthetic, transporter, transcriptional,
CC translational, signal transduction, DNA replication and cell division
CC activity. The method is useful for identifying a compound having the
CC ability to inhibit growth or proliferation of C. albicans cells and for
CC treating infection by C. albicans. The present sequence is that of an
CC essential Candida albicans protein used in the method of the invention.
CC Note: The sequence data for this patent is not represented in the printed
CC specification but is based on sequence information supplied to Derwent by
CC the European Patent Office
XX
XX Sequence 1881 AA;
Query Match 9.4%; Score 253.5; DB 5; Length 1881;
Best Local Similarity 21.6%; Pred. No. 5.4e-07;
Matches 100; Conservative 94; Mismatches 195; Indels 73; Gaps 12;
QY 52 LRDVSEELSRQLEDJLST-----YCVNNGGPGEDGAGCEPAEPEDAERSRYTVARNGP 107
Db 776 LEDKEQELAKIQEDHKSINKEKLVANSLOGIKARTKSETISGPDQQLQALKKGNTS 835
QY 108 EPTPVYGEKPSKGPNTTEIROSDVGDGRDHR-----PQEKKAKG 151
Db 836 ESTLKQLEKLD-----TEQAKKLEDQINNMTDLPFLKKSSEAEQIKQREPEPKN 890
QY 152 L-----CKEITLMQTLNTLSTP-EKLAALCKKYAEELLEHEHNSQOMK----- 195
Db 891 LTFEFTKKOYLOINLNKSNBPKINELSKISLSTEDNKFNAKQLEKLDTEB 950
QY 196 ----LLQKQSOVLQVQKHRCGE---HSAVLARSKLSESLCRLQHNRSLSKEGVQRAR 248
Db 951 NNEHLMQKRSASVAYNDLKAKASESEETVKAKELETLTLSKIDNLEKELKEQ-QSKKN 1009
QY 249 EEEERKQVTSHPQVTLNDIQLOMEQHNERNSKLQEMELAEKRLKLEIQLVELEEHID 308
Db 1010 ELEGQLQNTIDSTNEKFKLEDELSIKKSNSELSQNSQSELQKLEKTKQLQAKDEID 1069
QY 309 KYFKEKDLQOOLVDAKLQAOEMLKEAEERHOREKDFLLKEAVESORMCMLKQOETHLK 368
Db 1070 KLAETKSNIDNLNSEISSLSQSLKEAESHSSTVD---EHSLSLENLKKLEEVNTKT 1126
QY 369 QQLALYTEKPEEPQNTLSKSESVFTTFKQEMERMTKKIKLEKETTYRVSWESENKALL 428
Db 1127 SMIAKLSAKIEBHKA-----TDEIETTKTHITDLOBEHAKQKQSFESRNDIK 1175
QY 429 ENAEKTVRDKELEGQVQIQRKLQALQTERNDLANKRVQ 470
Db 1176 SNLDE---ANKELSDNREKLSNLEK-----EXTELANKLK 1207

RESULT 38
ADD46507
ID ADD46507 standard; protein; 1206 AA.
XX AC ADD46507;
XX 29-JAN-2004 (first entry)
XX Rat Protein BAA20077, SEQ ID NO 12188.
XX Rat; pain; neuronal tissue; gene therapy; spinal segmental nerve injury;
XX Chronic constriction injury; CCI; spared nerve injury; SNI; Chung.
XX Rattus norvegicus.
XX W02003016475-A2.
XX 27-FEB-2003.
XX 14-AUG-2002; 2002WO-US025765.
XX 14-AUG-2001; 2001US-0312147P.

PR 01-NOV-2001; 2001US-0346382P.
 PR 26-NOV-2001; 2001US-0333347P.
 PA (GHEO) GEN HOSPITAL CORP.
 PA (FARB) BAYER AG.
 PI Woolf C, D'urso D, Befort K, Costigan M;
 DR WPI; 2003-268312/26.
 DR GENBANK; BAA20077.
 XX New composition comprising two or more isolated polypeptides, useful for
 PT preparing a medicament for treating pain in an animal.
 XX Claim 1; Page; 1017pp; English.
 CC The invention discloses a composition comprising two or more isolated rat
 CC or human polynucleotides or a polynucleotide which represents a fragment,
 CC derivative or allelic variation of the nucleic acid sequence. Also
 CC claimed are a vector comprising the novel polynucleotide, a host cell
 CC comprising the vector, a method for identifying a nucleotide sequence
 CC which is differentially regulated in an animal subjected to pain and a
 CC kit to perform the method, an array, a method for identifying an agent
 CC that increases or decreases the expression of the polynucleotide sequence
 CC that is differentially expressed in neuronal tissue of a first animal
 CC subjected to pain, a method for identifying a compound which regulates
 CC the expression of a polynucleotide sequence which is differentially
 CC expressed in an animal subjected to pain, a method for identifying a
 CC compound that regulates the activity of one or more of the
 CC polynucleotides, a method for producing a pharmaceutical composition, a
 CC method for identifying a compound or small molecule that regulates the
 CC activity in an animal of one or more of the polypeptides given in the
 CC pain and a pharmaceutical composition comprising the one or more
 CC polypeptides or their antibodies. The polynucleotide or the compound that
 CC modulates its activity is useful for preparing a medicament for treating
 CC pain (e.g. spinal segmental nerve injury (SNI)), chronic constriction
 CC injury (CCI) and spared nerve injury (SNI) in an animal (e.g. gene
 CC therapy). The sequence presented is a rat protein (shown in Table 2 of
 CC the specification) which is differentially expressed during pain. Note:
 CC The sequence data for this patent did not form part of the printed
 CC specification, but was obtained in electronic form directly from WIPO at
 CC ftp.wipo.int/pub/published_pct_sequences.
 XX Sequence 1206 AA;
 SQ
 QY 12 EGACERPSQAAPAVEAGGSSQA-----PR 37
 DB 682 ENAQLPVKAEP-----CAPAASQASPPVLPISINHSNTENKMGALPPEITLPP 737
 QY 38 KPEGAQARTASG---ALRDVSEBLSQLEDILSTYCVNNQGGPGEDGAGGEPAPEDA 94
 DB 738 EPENGKGNDDTSGGTVSNESSDLNLSISFSL-----KTKDSQSVSLQETRRQKTL 791
 QY 95 EKSTYVARNGEPEPTPVYGEKPSKDPNTHIROSDEVDGRDHRPPOEKKKAGLKG 154
 DB 792 KTKRFIVDGVESVTT-----SKIVDSDSKTBELR-----FLRRQLRELRLIQR 838
 QY 155 EITLMLQTLN-TLSTPEKLAALCKYAELEHNSQKMKLIQKQSLVQEKDHLRG 213
 DB 839 BEQAQAQQLNGKLQOQREQIFRPEQ--EMLSKKQYDQBIENLEKQKQTIIE----- 889
 QY 214 EHSKAVLARSKLSLCELQHNHNSLKEGVQVAREEERKKEVTSFQVTLNDIQLQME 273
 DB 890 -----KLE-----QHTNRLADE-AKRIKGEQKE-----LSKFNMLRN----- 923
 QY 274 QHNERNKLRQENMLAEKLIQEQ-----YELREH 306
 DB 924 RKKEBEFVQKQOQLOGLAKKIIQQQKAEIANIERECLNNKQQLLAREAAIWELERH 983

QY 307 IDKVFVGHKDIQQQLVDAKIQAOEMLKEAE---ERHOREKDFLLKEAV---ESQMCMLMK 361
 DB 984 LQE--KHQLLKQOLKDDQYFIQRHQLLKRHEKTEQCRYNQRLIBELKRNQTOERARLPK 1041
 QY 362 QQETHLKQOLALY-----TEKFEFQNTLSKSSEVFTTTFKQEMEKMTKKI 406
 DB 1042 IQRSEAKTRWAMFKKSLRINSTATPDQREKIKQFAAQBEK-----RQKNERMAQH- 1092
 QY 407 KLEKETTMYRWSSENKALLEMABEK-----TVRDKLE---GLQVKLQRLKLC 455
 DB 1093 QKHESOMRDLQOQCEANVRELHQLQNEKCHLLVHEHTQKLKELDEHSELKEWR-EKLR 1151
 QY 456 RALQTERNDLNKRVQDLS-----AGGQSLTDSGPPRRPEGQAQAPSSPRVTEACPYPG 510
 DB 1152 PRKTLLEEFARKLQOEVEFFKMTGESECLNFSQSR-----GCLQTSHPSSTRAPAWAG 1206

RESULT 39
 ADD48875
 ID ADD48875 standard; protein; 1206 AA.
 XX AC ADD48875;
 XX DT 29-JAN-2004 (first entry)
 XX DE Rat Protein NP_062222, SEQ ID NO 14586.
 XX KW Rat; pain; neuronal tissue; gene therapy; spinal segmental nerve injury;
 XX chronic constriction injury; CCI; spared nerve injury; SNI; Chung.
 XX OS Rattus norvegicus.
 XX PN WO2003016475-A2.
 XX PD 27-FEB-2003.
 XX PF 14-AUG-2002; 2002WO-US025765.
 XX PR 14-AUG-2001; 2001US-0312147P.
 PR 01-NOV-2001; 2001US-0346382P.
 PR 26-NOV-2001; 2001US-0333347P.
 XX (GHEO) GEN HOSPITAL CORP.
 XX (FARB) BAYER AG.
 PI Woolf C, D'urso D, Befort K, Costigan M;
 XX WPI; 2003-268312/26.
 DR GENBANK; NP_062222.
 XX New composition comprising two or more isolated polypeptides, useful for
 PT preparing a medicament for treating pain in an animal.
 XX Claim 1; Page; 1017pp; English.
 CC The invention discloses a composition comprising two or more isolated rat
 CC or human polynucleotides or a polynucleotide which represents a fragment,
 CC derivative or allelic variation of the nucleic acid sequence. Also
 CC claimed are a vector comprising the novel polynucleotide, a host cell
 CC comprising the vector, a method for identifying a nucleotide sequence
 CC which is differentially regulated in an animal subjected to pain and a
 CC kit to perform the method, an array, a method for identifying an agent
 CC that increases or decreases the expression of the polynucleotide sequence
 CC that is differentially expressed in neuronal tissue of a first animal
 CC subjected to pain, a method for identifying a compound which regulates
 CC the expression of a polynucleotide sequence which is differentially
 CC expressed in an animal subjected to pain, a method for identifying a
 CC compound that regulates the activity of one or more of the
 CC polynucleotides, a method for producing a pharmaceutical composition, a
 CC method for identifying a compound or small molecule that regulates the
 CC activity in an animal of one or more of the polypeptides given in the
 CC pain and a pharmaceutical composition comprising the one or more
 CC polypeptides or their antibodies. The polynucleotide or the compound that
 CC modulates its activity is useful for preparing a medicament for treating
 CC pain (e.g. spinal segmental nerve injury (SNI)), chronic constriction
 CC injury (CCI) and spared nerve injury (SNI) in an animal (e.g. gene
 CC therapy). The sequence presented is a rat protein (shown in Table 2 of
 CC the specification) which is differentially expressed during pain. Note:
 CC The sequence data for this patent did not form part of the printed
 CC specification, but was obtained in electronic form directly from WIPO at
 CC ftp.wipo.int/pub/published_pct_sequences.
 XX Sequence 1206 AA;
 SQ

CC pain and a pharmaceutical composition comprising the one or more
CC polypeptides or their antibodies. The polynucleotide or the compound that
CC modulates its activity is useful for preparing a medicament for treating
CC pain (e.g. spinal segmental nerve injury (Chung), chronic constriction
CC injury (CCI) and spared nerve injury (SNI)) in an animal (e.g. gene
CC therapy). The sequence presented is a rat protein (shown in Table 2 of
CC the specification) which is differentially expressed during pain. Note:
CC The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic form directly from WIPO at
CC ftp.wipo.int/pub/published_pct_sequences.

XX Sequence 1206 AA;

Query Match 9.4%; Score 253; DB 7; Length 1206;
Best Local Similarity 22.2%; Pred. No. 3.4e-07;
Matches 133; Conservative 92; Mismatches 199; Indels 176; Gaps 25;

QY 12 EGAQERPSQAAPAVEAGPSSOA-----PR 37
DB 682 ENAQELFVKAEP-----QAPAAQSAPPPVLPISNIHSENTKNGEMALPKPETILPP 737
QY 38 KPFGAQAARTAQSG---ALRDVSEELSRQLEDILSTYCVNNQGGPGEDGAQGEPAEPEDA 94
DB 738 EPENGKNDTSGTGVNENSSDNLSTSSFLS-----KTKDSGVSLQETRRQKTL 791
QY 95 EKRTYVARGPEPTPVVYGEKPKGDNTEIRQSDVEVGDRHRRPQEKKAQGLQK 154
DB 792 KTKRKPTVDGVEVSVTT---SKIVTSDSDSKTEELR-----FLRQELRELLQK 838
QY 155 EITLLMOTLN-TLSTPEEKLAALCKYAELEHRNSQOMKLLQKQSQOLVQEKDHLRG 213
DB 839 ESKAQOOLNGKLOQREQIFRPEQ--EMLSKQYDQEIENLERQKQQTIE-----889
QY 214 EHSKAVLARSKLBSLRELQHRNSLKBEQVQAREREKKEVTSHFQVTLNDIQLQWE 273
DB 890 -----RLS-----QETNRLRDE-AKRIKGEQKE---LSKFQNMRLN-----923
QY 274 QHNRNSKLQENMELAEKLLLEO-----YELREH 306
DB 924 RKKEBEFVQKQOQLDGLAKLIIQOKQAEIANIERCLNNKQOOLARARAEALWELEERH 983
QY 307 IDKVFHKDILQOQLVDAKLAQQAQEMLEAE--ERHQREKDFLLKEAV--ESQRMCELMK 361
DB 984 LQB--KHQLLQKLDQYFQORHQLKREKETEQQRYNORLIBELKNRQTOERARLPK 1041
QY 362 QQTHLKQOALY-----TEKFEFQNTLSKSSEVTTTFQEMERKTKI 406
DB 1042 IQRSEAKTRMAMFKSLRINSTATPDQDREKIKQFAAQEEK-----RQKNERMAQH- 1092
QY 407 KLEKETTYRWSWSSNALLMALEK-----TVRDKLE--GLQVKIQRLEKLC 455
DB 1093 QKHESQWRDLQACEANVRELHQIQNEKCHLLVEHETQKLEDEHSQELKEWR-EKLR 1151
QY 456 RALQTERNDLNKRVQDLS-----AGGGSLTDSGPRRPGGQAQAPSPRVTEAPCPYG 510
DB 1152 PRKKTLEEPARKLQEQVEFFQMTGSECLNPSAQSR-----GCLQTSHPSSSTRAPAWAG 1206

RESULT 40
AAB96332
ID AAB96332 standard; protein; 880 AA.
XX AC AAB96332;
XX DT 29-OCT-2001 (first entry)
XX DE Putative P. abyssi ATPase involved in DNA repair #2.
XX KW Hyperthermophilic archaeon; hyperthermophilic protein.
XX OS Pyrococcus abyssi.
XX PR 292651-A1.

XX 27-OCT-2000.
XX PD
XX 21-APR-1999; 99FR-00005034.
XX PR 21-APR-1999; 99FR-00005034.
XX PA (CNRS) CNRS CENT NAT RECH SCI.
XX PA (IPRE-) IPREMER INST FR RECH EXPL MER.
XX PI Porterre P, Thierry JC, Prieur D, Dietrich J, Lecompte O;
XX PI Querellou J, Weissenbach J, Saurin W, Hellig R;
XX WP1; 2001-126236/14.
XX New nucleotide sequences isolated from Pyrococcus abyssi encode proteins
XX useful in industry.
XX Claim 7; Page 1003-1006; 1657pp; French.
XX The present invention relates to the genomic sequence of Pyrococcus
XX abyssi (see AAF8431 and AAH41223-7) and P. abyssi proteins. P. abyssi is
XX a hyperthermophilic archaeon, which is isolated from deep-sea
XX hydrothermal vents. The present sequence is one such P. abyssi protein.
XX The proteins of the present invention have various potential industrial
XX uses, since the proteins are stable at very high temperatures, some up to
XX 110 degrees centigrade. Note: This patent is in the same patent family as
XX WO200005062, which contains additional sequences as shown in AAB99132-
XX AAB99143, AAH75903-AAH75920 and AAG66436
XX Sequence 880 AA;

Query Match 9.3%; Score 252.5; DB 4; Length 880;
Best Local Similarity 23.2%; Pred. No. 2.6e-07;
Matches 123; Conservative 98; Mismatches 183; Indels 127; Gaps 22;

QY 12 EGAQERPSQAAPAVEAGPSSQ-----APRKEGAQAARTAQSGALRDVSEELSRQLE 64
DB 259 KGLEEKIYQIERSIEBKAKISELBEIVKDIPLQKEKEKYLKAG-FRDEYESKLRLLE 317
QY 65 DILSTYCVNNQGGPGEDGAQGEPAEPDAKSRTRYVARGPEPTPVVYGEKPKSGDP 124
DB 318 KELSFW-----ESLKAIBEVIKEGE-----KKKB 342
QY 125 NTEEIROS-DEVGDR-DHRRP--QEKKAAGLKGKITLLMOTLNTLSTPE--EKLAAALCK 178
DB 343 RABEIREKLSERKELELKPVELEDAKQVQKQIERLKARLKLSPGEVTEKLESLEK 402
QY 179 KYAELEHRNSQOMKLLQKQSQOLVQEKDHLRGEHSKAVLARSKLSLC-----REL-Q 233
DB 403 ERTEI-----EBAKEITTRIGOMEQKN-----ERKAELEELKAKGKPCVCGRELTE 451
QY 234 RHNRSLK-----EGVQARABEEKKEVTSHFQVTLNDIQLQ---MEQHNRNS 280
DB 452 EHKELMERYTLIEIKIEELKRTTEBRKURVNLKLEIKLUREFSVMDIAEQIKELBS 511
QY 281 KLQRQNMELAE-----RLKKGLEIQLYELREHIDKVPK 312
DB 512 KLKGFNLEELQKEREPEGLNEEFNKLKGLGLERDLKRIKALEGRKKLIEKVRKAKE 571
QY 313 H-KLQOQQLVDAKLAQQAQEM--LKAEEHOREKDFLLKEAVESQRMCELMKQETHLK 368
DB 572 ELENLHRLQRLGPFSEVEELNLRIOELEEFHDKYVAAKXSES-ELRELKMKLEKSKTELD 630
QY 369 QQLALYTKFBEFQNTLSKSSEVFTTF--QKEMKMTKKIKLEKETTYRWSWSSNKA- 426
DB 631 QAFEMLADVENIEEKEAKLDLESKFNEEYEEKERLVKLEREVSSILARLEELKKS 690
QY 427 -----LLEMAFEKTVRKLEGLQVKTORLEKLCRALQTERNDLNKRVQD 471
DB 691 EQKATLAKLKEEKEREK---AKLEIKKLEKALSKEV---DLRKKIKD 733

```
RESULT 41
ABB61144
ID ABB61144 standard; protein; 1690 AA.
XX
AC ABB61144;
XX
XX 26-MAR-2002 (first entry)
XX
DE Drosophila melanogaster polypeptide SEQ ID NO 10224.
XX
XX Drosophila melanogaster polypeptide; cell signalling; insecticide;
XX
XX Drosophila melanogaster.
XX
XX WO200171042-A2.
XX
XX 27-SEP-2001.
XX
XX 23-MAR-2001; 2001WO-US009231.
XX
XX 23-MAR-2000; 2000US-0191637P.
XX
XX 11-JUL-2000; 2000US-00614150.
XX
XX (PEKE ) PE CORP NY.
XX
XX Venter JC, Adams M, Li PWD, Myers EW;
XX
XX WPI; 2001-656860/75.
XX
XX N-PSDB; ABL05247.
XX
XX New isolated nucleic acid detection reagent for detecting 1000 or more
XX genes from Drosophila and for elucidating cell signalling and cell-cell
XX interactions.
XX
XX Disclosure; SEQ ID NO 10224; 21pp + Sequence Listing; English.
XX
XX The invention relates to an isolated nucleic acid detection reagent
XX capable of detecting 1000 or more genes from Drosophila. The invention is
XX useful in developmental biology and in elucidating cell signalling and
XX cell-cell interactions in higher eukaryotes for the development of
XX insecticides, therapeutics and pharmaceutical drugs. The invention
XX discloses genomic DNA sequences (AB16176-ABL30511), expressed DNA
XX sequences (ABL01840-ABL16175) and the encoded proteins (ABB57737-
XX ABB72072). The sequence data for this patent did not form part of the
XX printed specification, but was obtained in electronic format directly
XX from WIPO at ftp.wipo.int/pub/published_pct_sequences
XX
XX Sequence 1690 AA;
XX
XX Query Match 9.3%; Score 251.5; DB 4; Length 1690;
XX Best Local Similarity 23.3%; Pred. No. 6.3e-07;
XX Matches 127; Conservative 102; Mismatches 180; Indels 135; Gaps 23;
XX
XX 33 SQAPKPGGAQRTAQSGALDVSELSRQLEDILSTYCVNNQCGFGDGAQGEPAEPE 92
XX 352 SSTPVKPLATPKSQFS--MODLREKQQRVKELMVERDLDR-----E 392
XX
XX 93 DAESRTYVARNGEPEPTPVYVYGEKPSKDPNTEIRQS--DEV---GDRDHRPQ--EXK 147
XX
XX 393 DAQWALQLOXNINELKXRIYELASALDNERRKTEELQCSIDEAQFCGDELNAQSQVYKE 452
XX
XX 148 KAKGKGKEITLMO---TINTLSTP-----EKLALCKK-----YALL 184
XX
XX 453 KIHDLSEKITLVATSPSLQSLPDLPSDDGALQEEIAKLQKMTIQKEVESRIAEOL 512
XX
XX 185 EERNRSOKMLLOK---OSQVQEKDHLRGHSHKAVLARSKLSELCRELQRHNSLK 240
XX
XX 513 EERQLRNVKVLNEQIATLQSELVSKDEAL-----EKFSLECCQENLRLE---LLX 564
XX
XX 241 EGVQRARER-----EKQKEV-----TSHPOVTLNDI 268
XX
XX 565 ENEKQAEQAEFTKLAERKSEVILRLSELQNLKATSDSLESERVNKTDECEILQTEV 624
```

Query Match 9.3%; Score 251.5; DB 4; Length 1690;

XX DT 18-FEB-2000 (first entry)
 XX DE Human SULU3 protein.
 XX KW Antirheumatic; antiarthritic; antinflammatory; antiallergic; osteopathic;
 KW antipsoriatic; antiarteriosclerotic; antiasthmatic; immunosuppressive;
 KW neuroprotective; cardiant; cerebroprotective; cytostatic; antidiabetic;
 KW vulnery; STE20; protein kinase; STLK2; STLK3; STLK4; STLK5; STLK6; STLK7;
 KW ZC1; ZC2; ZC3; ZC4; KHS2; SULU3; SULU4; SULU5; STLK2; STLK3; STLK4; STLK5; STLK6; STLK7;
 KW inflammatory bowel disease; Crohn's disease; osteoarthritis; psoriasis;
 KW rhinitis; autoimmunity; organ transplantation; multiple sclerosis;
 KW myocardial infarction; cardiovascular disease; stroke; renal failure;
 KW oxidative stress-related neurodegenerative disorder; Parkinson's disease;
 KW amyotrophic lateral sclerosis; Leigh syndrome; cancer; cardiomyopathy;
 KW ischemic disorder; inflammation; diabetes mellitus; fibrosis; mitosis;
 KW mesangial disorder; growth regulation; wound healing; T cell activation;
 KW immunosuppressant.
 XX OS Homo sapiens.
 XX PN WO9953036-A2.
 XX XX 21-OCT-1999.
 XX PD 13-APR-1999; 99WO-US008150.
 XX PF 14-APR-1998; 98US-0081784P.
 XX PR (SUGE-) SUGEN INC.
 XX PA Plowman G, Martinez R, Whyte D;
 XX PI WPI; 1999-611301/52.
 XX DR N-PSDB; AAZ40489.
 XX DT Novel kinase-related polypeptides used for the diagnosis and treatment of
 DT kinase-related diseases and disorders.
 XX Claim 11; Page 296-299; 387pp; English.
 XX This sequence represents a novel STE20-related protein kinase. The
 CC invention relates to nucleic acid molecule encoding a kinase polypeptide
 CC selected from STLK2, STLK3, STLK4, STLK5, STLK6, STLK7, ZC1, ZC2, ZC3,
 CC ZC4, KHS2, SULU1, SULU3, SULU4, PAK4 and PAK5. The proteins are used to
 CC identify agonists and antagonists, and to raise antibodies. The
 CC polynucleotides are useful in gene therapy protocols. The polynucleotides,
 CC polypeptides, antibodies, antagonists and agonists may be used to treat
 CC diseases such as immune-related disorders and diseases (e.g. rheumatoid
 CC arthritis, atherosclerosis, chronic inflammatory bowel disease (e.g.
 CC Crohn's disease), asthma, osteoarthritis, psoriasis, atherosclerosis,
 CC rhinitis, autoimmunity, and organ transplantation, chronic inflammatory
 CC pelvic disease, multiple sclerosis, organ transplantation, myocardial
 CC infarction, cardiovascular disease, stroke, renal failure, oxidative
 CC stress-related neurodegenerative disorders (e.g. amyotrophic lateral
 CC sclerosis, Parkinson's disease and Leigh syndrome), cancer,
 CC cardiomyopathies, ischemic disorders, inflammatory disorders, diabetes
 CC mellitus, fibrotic and mesangial disorders. The proteins may also be
 CC useful for cell growth regulation (e.g. in wound healing), T cell
 CC activation, mitosis control, and as immunosuppressants
 XX SQ Sequence 786 AA;
 XX Query Match 9.2%; Score 248.5; DB 2; Length 786;
 XX Best Local Similarity 21.6%; Pred. No. 4e-07;
 XX Matches 139; Conservative 117; Mismatches 197; Indels 191; Gaps 29;
 XX QY 23 PAVEAE-----GFGSSQA-PRKEGAQAFTAGSALRDVSELSQLE 64
 XX D0 109 PAVEAQEEEDHGVGRTGTGVNSVGSNQSTPSMSISASSQSSSVNSLDPVSDKKS-ELD 167
 XX QY 65 DILSTYCVNNQ-----GFGEDG---AQGPAPEDAEKSRITYVANGEPPTVPVYGEK 118

Db 168 MEGDHTVMSNSVHLKPEENYREEDP-----RT---RASDPQSPQVSRHKS 215
 QY 119 PSKGPNTBEIRQSDVGRDRHRRPQEKCAKGLQKEIT-----LLMQTNTLTSTPBKLA 174
 Db 216 HYRNRHPATIRTSILV-----TRQMEHQDSLEAQSGYKRRRQHQKQLATLENK 271
 QY 175 ALCKYA-----ELLEHRN--SOKWKLQKQSOINVQ-----KDLRGEHSHK 217
 Db 272 AEWDEHRLDKOLETQORNFANAEKLIKQAAKEKAKVMSNEEKFKQHQIQQQKK 331
 QY 218 AVLAHSKLESCLRELQRHNRSLKBEQVQ-----ARBEER---KRKEVTSHPQV--TLNDI 268
 Db 332 EL--NSFLSQKREYKLRKEQLKEELNENQSTPKKEQSWLSKQENIQHQAEBEANLL 389
 QY 269 QLOME-----CHN-----ERNKLEQENMELAEARLKLILQVELRE 304
 Db 390 RHQRQVLELCRRFKRMMLGRHNLQDLVRELAKRQTKDLEHMLLRQSHESQOELEF 449
 QY 305 EHIDKVFHKDLQQLVDLAKQAQEMLKEAEERHQREKDFLLKEAVESQRCMLKQOE 364
 Db 450 RHINTIQK---MRCELI--RLQHOTELTNQLEYNKRRERELRRKRVMEVQCPKLSKSE 504
 QY 365 THLKQOL-----ALYTKPEEFONTLSKSEVTTTQKQEMKMTKKIKLEKETTMYRSW 420
 Db 505 LQIKQFQDTCKIQRYKALNHLLE-----TTPKSEHKAVLRLK--EETRLKLA 557
 QY 421 ESNKALLEMAEKTVR-----DKELE 442
 Db 558 EYDHSINEMLSQALRLDEAQAECQVLMKQQLQLELLNAYQSKIKNQASQAQHDREL 617
 QY 443 GLQVKI-----QRLEKLCRALQTERND-----LNRVQDLASAGQGS--L 480
 Db 618 ELEQRVSLRRLLEQKIEEMALQNERTERISLERQARIEAFDSSEMLGFSMMVL 677
 QY 481 TDSGPERREPGQAQAPSSPRVTEAPCYGAP--STEASGQTGP 522
 Db 678 SNLSPE-----AFSHSYFGASGWSHNPTGGP 705
 RESULT 45
 AAV55942
 ID AAV55942 standard; peptide; 1001 AA.
 AC AAV55942;
 XX DT 19-FEB-2000 (first entry)
 XX Human/Murine SULU3 consensus protein sequence.
 KW Antirheumatic; antiarthritic; antinflammatory; antiallergic; osteopathic;
 KW antipsoriatic; antiarteriosclerotic; antiasthmatic; immunosuppressive;
 KW neuroprotective; cardiant; cerebroprotective; cytostatic; antidiabetic;
 KW vulnery; STE20; protein kinase; STLK2; STLK3; STLK4; STLK5; STLK6; STLK7;
 KW ZC1; ZC2; ZC3; ZC4; KHS2; SULU3; SULU4; SULU5; STLK2; STLK3; STLK4; STLK5; STLK6; STLK7;
 KW inflammatory bowel disease; Crohn's disease; osteoarthritis; psoriasis;
 KW rhinitis; autoimmunity; organ transplantation; multiple sclerosis;
 KW myocardial infarction; cardiovascular disease; stroke; renal failure;
 KW oxidative stress-related neurodegenerative disorder; Parkinson's disease;
 KW amyotrophic lateral sclerosis; Leigh syndrome; cancer; cardiomyopathy;
 KW ischemic disorder; inflammation; diabetes mellitus; fibrosis; mitosis;
 KW mesangial disorder; growth regulation; wound healing; T cell activation;
 KW immunosuppressant.
 XX OS Homo sapiens.
 XX PN WO9953036-A2.
 XX XX 21-OCT-1999.
 XX PD 13-APR-1999; 99WO-US008150.

XX 14-APR-1998; 98US-0081784P.
XX (SUGS-) SUGEN INC.
XX Plowman G, Martinez R, Whyte D;
XX WPI; 1999-611301/52.
XX Novel kinase-related polypeptides used for the diagnosis and treatment of
XX kinase-related diseases and disorders.
XX Claim 11; Page 312-315; 387pp; English.
XX
XX This sequence represents a consensus peptide sequence contained in novel
XX STE20-related protein kinases. The invention relates to a nucleic acid
XX molecule encoding a kinase polypeptide selected from STK2, STK3, STK4,
XX STK5, STK6, STK7, ZC1, ZC2, ZC3, ZC4, ZC5, SULU1, GEX2, PAK4
XX and PAK5. The proteins are used to identify agonists and antagonists, and
XX to raise antibodies. The polynucleotides are useful in gene therapy
XX protocols. The polynucleotides, polypeptides, antibodies, antagonists and
XX agonists may be used to treat diseases such as immune-related disorders
XX and diseases (e.g. rheumatoid arthritis, arteriosclerosis, chronic
XX inflammatory bowel disease (e.g. Crohn's disease), asthma,
XX osteoarthritis, psoriasis, atherosclerosis, rhinitis, autoimmunity, and
XX organ transplantation, chronic inflammatory pelvic disease, multiple
XX sclerosis, organ transplantation, myocardial infarction, cardiovascular
XX disease, stroke, renal failure, oxidative stress-related
XX neurodegenerative disorders (e.g. amyotrophic lateral sclerosis,
XX Parkinson's disease and Leigh syndrome), cancer, cardiomyopathies,
XX ischemic disorders, inflammatory disorders, diabetes mellitus, fibrotic
XX and mesangial disorders. The proteins may also be useful for cell growth
XX regulation (e.g. in wound healing), T cell activation, mitosis control,
XX and as immunosuppressants
XX
XX Sequence 1001 AA;
XX
XX Query Match 9.2%; Score 248.5; DB 2; Length 1001;
XX Best Local Similarity 21.6%; Pred. No. 5.3e-07;
XX Matches 139; Conservative 117; Mismatches 197; Indels 191; Gaps 29;
XX
XX 23 PAVRAB-----GPGSSQA-PRKPECAQARTAQSGALRDVSEELSGLE 64
XX 324 PAVEAQEEBEEBQHGVRGTGVNSQNSIPMSISASSQSSVNSLPDVSDDKS-ELD 382
XX 65 DILSTYCVNNQG-----GPGEDG--AQGEPAEPDAEKSRITYVARNGBEPTFVYGEKE 118
XX 383 MMEGDHVMNSVVIHLKPEENVREGDP-----RT---RASDPQSPQVSRHKS 430
XX 119 PSKGDPTNTEIROSDVGDHRRPQRKKAQGLGKEIT-----LLMOTLNTLSTPEKLA 174
XX 431 HYRNREHFATIRTSALV---TRQMGEHQDSLELREQMSGYKMRMRQHOKQLMTLENK 486
XX 175 ALCKKVA---ELLEERN--SOKMKLLQKDSOLVOE-----KDLHGEHKS 217
XX 487 AEMDEHLRLDKDLETQRNFAEMKLIKQHAAMEKEAKVMSNEKKFQHIQAOQK 546
XX 218 AVIARSKLESCLRELQHNLSKRGVQR-----AREEEE---KKRVTGHFQV--TINDI 268
XX 547 EL--NSFLSQKREYKLRQLKEELNENQSTPKCKQEWLSKQENIQHFOAEENALL 604
XX 269 QLOWE-----QNN-----ERNKLRQENNELAERLKLIEQYELAE 304
XX 605 RRQRYLELESCRFKPKMLGRNLEODLVRELNKRQTKOLEHAMLRLQHESMGELEP 664
XX 305 EHKDKVPMKDLQQLQVADAKLQQAQENLKAEBERHOREKDFLKEAVESQRMCELMKQOE 364
XX 665 RHLNTIQK---MRCELI--RLQHTQELTNLQYNNKRRERELRKHVMYVRQPKSLSKS 719
XX 365 THLKKOQL-----ALYTEKFEFPQNTLSKSEVFTTFQEMEKMTKKIKLEKETMYTRSW 420
XX 720 LQIKQFQDQCKTQTRQYKALRNHLE-----TTPKSEHKAVLKLK--BEQTKRLAILA 772

QY 421 ESNKALLEMAEKEKTVR-----DKLELE 442
Db 773 EYVDHSINEMLSQALRLDEAQEAECQVLKMQQLQLELLNAYQSKIKMQAEQHDREL 832
QY 443 GLQVKI-----QLEKLCALOTERND-----LNKRVODLSAGQGS 480
Db 833 ELQQRVSLRRALLEQKIEEMALQKQERIRSLRQAREIEAFDSMRGLGFSNMVL 892
QY 481 TDSGPERPEGCGAQAQSPSPRVTEAPCYGAP--STEASGQTGP 522
Db 893 SNLSPE-----AFSHSYDYGASGWSHNPFGPGP 920

RESULT 46

AB897326
ID ABB97326 standard; protein; 1001 AA.

XX ABB97326;

XX 27-JUN-2002 (first entry)

XX Novel human protein SEQ ID NO: 594.

XX Human; antianaemic; vulnary; antiinflammatory; immunomodulator;
XX antifertility; cerebroprotective; cytostatic; rheumatic; gene therapy;
XX neuroprotective; antiparkinsonian; protein therapy; EST;
XX expressed sequence tag.

XX Homo sapiens.

XX WO200222660-A2.

XX 21-MAR-2002.

XX 10-SEP-2001; 2001WO-US026015.

XX 11-SEP-2000; 2000US-00659671.

XX (HYSE-) HYSEQ INC.

XX Tang YT, Liu C, Zhou P, Asundi V, Zhang J, Zhao QA, Ren F;
XX Xue AJ, Yang Y, Wehrman T, Drmanac RT;

XX WPI; 2002-292408/33.

XX N-PSDB; ABN32512.

XX An isolated polynucleotide for treating diseases associated with its
XX encoded polypeptide such as cancer and multiple sclerosis.

XX Example 2; SEQ ID NO 594; 509pp; English.

XX The present invention provides the protein and coding sequences of 444
XX novel human proteins. These are isolated from expressed sequences tags
XX (ESTs). They can be used to stimulate cell growth, to regulate
XX haematopoiesis e.g. to treat aplastic anaemia, to help tissue regrowth
XX e.g. in burn treatment, to regulate the immune system e.g. to treat
XX multiple sclerosis, to regulate activin or inhibin e.g. to treat
XX infertility, to regulate haemostasis or thrombolysis e.g. to treat stroke
XX and cancer, to screen for drugs, to treat inflammatory conditions e.g.
XX rheumatoid arthritis, and to treat nervous system disorders e.g.
XX Parkinson's disease. The present sequence is a protein of the invention

XX Sequence 1001 AA;

Query Match 9.2%; Score 248.5; DB 5; Length 1001;

Best Local Similarity 21.6%; Pred. No. 5.3e-07;

Matches 139; Conservative 117; Mismatches 197; Indels 191; Gaps 29;

QY 23 PAVRAB-----GPGSSQA-PRKPECAQARTAQSGALRDVSEELSGLE 64

Db 324 PAVEAQEEBEEBQHGVRGTGVNSQNSIPMSISASSQSSVNSLPDVSDDKS-ELD 382

QY 65 DILSTYCVNNQG-----GPGEDG--AQGEPAEPDAEKSRITYVARNGBEPTFVYGEKE 118

Db	383	MMGEHTVMSSNVHLKPEBENTREGDP-----RT---RASDPSPQVSRHKS	430
Qy	119	PSKDPNTEIIRQSDVGRDRHRPQBKKAKGLGKEIT--LIMQTLNTLSTPEEKLA	174
Db	431	HYRNEHFATIRTSALV---TRQWQEHQDSELRQMSGYKMRHQHOKQLMTLENKLG	486
Qy	175	ALCKKYA---ELLEHRN--SOKMKLQKKQSOLVQ-----KQHLRGEHSK	217
Db	487	AMWDHRLRLDKDQTNNAFAEMKLIKKGQAMEKEAKVMNSBEKFPQOHIOAQQCK	546
Qy	218	AVLARSKLESLELQRHNRSLKEGVQR-----AREEEE---KKKEVTSHFQV--TLNDI	268
Db	547	EL--NSPLESQKREYKLRKEQLKEELNENQSTPKKEQEWLSKQENIQHFQAEBSANLL	604
Qy	269	QLQME-----OHN-----ERNKLRQENWELAEERLKKLIQVYELRE	304
Db	605	RRQRYLELCRRPKRMLGRHNLBDLVREELNRSQTKDLEHAWLLRQHESMQELEF	664
Qy	305	EHDIVFVKHDLQQQLVDAKLQQACQEWLMEKAEERHQRKDFLLKEAVESQRMCELMKQOE	364
Db	665	RHLNTIQK--MRCELI--RLHQHTELTNQLEYNKRERELRKHVNEVROQPKSLKSE	719
Qy	365	THLKOOL---ALYTEKEFEFPNTLSKSSVFITPKQEMKMTKKIKKLEKTTMTVRSRW	420
Db	720	LOIKKQFQDTCKIQTRQYKALRNHLLE-----TTPKSEHKAVLKRLK--EEQTRKLAILA	772
Qy	421	ESSNKALLEMAEEKTVR-----DKELE	442
Db	773	EQYDSHINELSTQALRLDEDAQBAQCVLRLMQQLELELNAVQSKIKQAEQAQHDREL	832
Qy	443	GLQVKI-----ORLEKLCRALQTERND-----LNKRVDQLSAGQGS-----L	480
Db	833	ELEQRVSLRALLEQKIEEMALQNERIIRSLERQARETEAPDSMSRLGFSNMVL	892
Qy	481	TTSGERRPREGQAAPSRTVAPCYPGAP--STEASQGTGP	522
Db	893	SNLSPE-----AFSHSYPGASGWSHNPTGGPGP	920
RESULT 47			
ADBS5356			
ID	ADBS5356	standard; protein; 1001 AA.	
XX	ADBS5356;		
XX	AC		
XX	DT		
XX	DE	Human Protein XP_030845, SEQ ID NO 1171.	
XX	DE		
KW	KW	Human; pain; neuronal tissue; gene therapy;	
KW	KW	spinal segmental nerve injury; chronic constriction injury; CCI;	
KW	KW	spared nerve injury; SNI; Chung.	
CS	CS	Homo sapiens.	
XX	XX		
FN	FN	WO2003016475-A2.	
XX	XX		
PD	PD	27-FEB-2003.	
XX	XX		
PF	PF	14-AUG-2002; 2002WO-US025765.	
XX	XX		
PR	PR	14-AUG-2001; 2001US-0312147P.	
PR	PR	01-NOV-2001; 2001US-0346382P.	
PR	PR	26-NOV-2001; 2001US-033347P.	
XX	XX		
PA	PA	(GEHO) GEN HOSPITAL CORP.	
PA	PA	(FARB) BAYER AG.	
XX	XX		
PI	PI	Woolf C, D'urso D, Befort K, Costigan M;	
XX	XX		
DR	DR	WPI; 2003-368312/26.	
DR	DR	GENBANK; XP_030845.	

XX
PT New composition comprising two or more isolated polypeptides, useful for
PT preparing a medicament for treating pain in an animal.
XX
PS
PS Claim 1; Page; 1017pp; English.
XX
XX The invention discloses a composition comprising two or more isolated rat
CC or human polynucleotides or a polynucleotide which represents a fragment,
CC derivative or allelic variation of the nucleic acid sequence. Also
CC claimed are a vector comprising the novel polynucleotide, a host cell
CC comprising the vector, a method for identifying a nucleotide sequence
CC which is differentially regulated in an animal subjected to pain and a
CC kit to perform the method, an array, a method for identifying an agent
CC that increases or decreases the expression of the polynucleotide sequence
CC that is differentially expressed in neuronal tissue of a first animal
CC subjected to pain, a method for identifying a compound which regulates
CC the expression of a polynucleotide sequence which is differentially
CC expressed in an animal subjected to pain, a method for identifying a
CC compound that regulates the activity of one or more of the
CC polynucleotides, a method for producing a pharmaceutical composition, a
CC method for identifying a compound or small molecule that regulates the
CC activity in an animal of one or more of the polypeptides given in the
CC specification, a method for identifying a compound useful in treating
CC pain and a pharmaceutical composition comprising the one or more
CC polypeptides or their antibodies. The polynucleotide or the compound that
CC mediates its activity is useful for preparing a medicament for treating
CC injury (CCI) and spared nerve injury (SNI) in an animal (e.g., gene
CC therapy). The sequence presented is a human protein (shown in Table 2 of
CC the specification) which is differentially expressed during pain. Note:
CC The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic form directly from WIPO at
CC ftp.wipo.int/pub/published pct sequences.

Sequence 1001 AA;

Query Match 9.2%; Score 248.5; DB 7; Length 1001;

Query Match	9.23; Score 258.5; DB 7; Range 1004;
Best Local Similarity	21.68; Pred. No. 5.3e-07;
Matches 139: Conservative	117; Mismatches 197;
Indels	191;
Gaps	29;

Qy	23	PAYEA-----GPOSSQ-PRKPEGAQARTQSGALRDVSELSRQLE	64
Db	324	PAYEAQEEBEQHGVRTVNSVGSNOIPMSISASSOSSVNSPLPDDSDKS--ELD	382
Qy	65	DILSTYCDVNNQG---GPCEDG--AQGSPAEPAEDAESKRTYYVARNGEPSTPVVYGEK	118
Db	383	MMEGDHTVMSNSSVIHLKPEEENYREGDP-----RT---RASDPQSPQVSRHKS	430
Qy	119	PSKGDNTBTEIROSDVGDDBDHRPRPKKAKGLGKEIT-----LLMOTLNTLSTPEKLA	174
Db	431	HYNRREHFATIRASLV---TRQMOHEODSBLREQMSGYKMRROHQOLMTLENKLK	486
Qy	175	ALCKKYA-----BLLEHRN--SOKQMKLKOKQSOLVOE-----KDLHRGBHSK	217
Db	487	AEMDSHRLDKDQLETQRNPPAAMEKLIKQHQAAMEKAYMSNEEKKTCQHILQAQCKK	546
Qy	218	AVLARSKLSCLRELQENRSLKEEGVQR---ARBEEE--KKXVTSHPOV--TLNDI	268
Db	547	EL--NSGFLESQKEVYKLKEQLKEELNENQSTPKGKQEWLSQKENIQHQAEEBANLL	604
Qy	269	QLOWE-----QHN-----RNSKLRQENNELAERLKKLIBQYELRE	304
Db	605	RRORQVYLECRFPKRMLLGRNLEBODLVREELNTRQTDLEHMLLRQHESMQLEFP	664
Qy	305	EHTDKVFKHXDLQOQLVDALQQAQEWMLKABERHQEKDPLIKEAVESQRMCMLKQOE	364
Db	665	RHLNTIQK--MRCELL--RLQHOTELTNQLEYNKGRERELRKRKHVNEVRQPKSLKSKE	719
Qy	365	THLKKQI-----ALYTEKPEFPNTLSKSSVEPTTFKQEMEKVKKIKKLKETTMYRSRW	420
Db	720	LQIKQFQDTCITQYKALRNHLLLE-----TTPSEHKAVILKRLK--SEOTKLAILA	772
Qy	421	ESSNKALLEMAEAKTVR-----DKELE	442

```

Db 773 EQYDHSINEMLSQALRLDEAQAECQVLKQQLQOELELLNAYQSKIKWCAQAQHDREL 832
QY 443 GLQVKI-----QLEKLCRALQTERND-----LNKRVODLSAGQGS-----L 480
Db 833 ELEQVSLRLAELEQKTEEMALQNERTERISLLERQARETEAFDSMRLGFSNMVL 892
QY 481 TDSGPERPBGQAQAPSSPRVTEAPCYGAP--STEASQOTGP 522
Db 893 SNLSPE-----AFSHSYFGASGWSHNPTGGPGP 920

RESULT 48
ABR47509
ID ABR47509 standard; protein; 1005 AA.
AC ABR47509;
XX
XX 12-JUN-2003 (first entry)
XX
XX Breast cancer associated protein sequence SEQ ID NO:254.
XX Human; breast cancer; cytostatic; gene therapy.
XX Homo sapiens.
XX W02003004989-A2.
XX
XX 16-JAN-2003.
XX
XX 21-JUN-2002; 2002WO-US019669.
XX
XX 21-JUN-2001; 2001US-0299887P.
XX 27-JUN-2001; 2001US-0301572P.
XX 18-JUL-2001; 2001US-0306501P.
XX 25-SEP-2001; 2001US-0325002P.
XX 05-MAR-2002; 2002US-0362585P.
XX 14-MAY-2002; 2002US-0380391P.
XX
XX (MILL-) MILLENIUM PHARM INC.
XX
XX Lillie J, Gannavarapu M, Giatt K, Hoersh S, Kamatkar S;
XX Mertens M, Monahan JE, Myer V, Xu Y, Zhao X, Meyers RE;
XX Bast RC, Hortobagyi GN, Fuzsai L, Meric P, Sahin A, Mills GB;
XX
XX WPI; 2003-210381/20.
XX N-PSDB; ACC50205.
XX
XX Breast cancer diagnosis or treatment by comparing the level of expression
XX of a marker in a patient sample with that in the control non-breast
XX cancer sample.
XX
XX Claim 1; SEQ ID NO 254; 128pp; English.
XX
XX The present invention describes a method for assessing whether a patient
XX is afflicted with breast cancer. The method comprises comparing the level
XX of expression of a marker (gene/polypeptide see ACC50076 to ACC50334 and
XX ABR47386 to ABR47632) in a patient sample and the normal level of
XX expression of the marker in a control non-breast cancer sample, where a
XX significant increase in the level of expression of the marker in the
XX patient sample and the normal level is an indication that the patient is
XX afflicted with breast cancer. The breast cancer associated sequences from
XX the present invention have cytostatic activities and can be used in gene
XX therapy. The method is useful for diagnosing and treating breast cancer.
XX N.B. The sequence data for this patent did not form part of the printed
XX specification, but was obtained in electronic format directly from WIPO
XX at ftp.wipo.int/pub/published_pct_sequences
XX
XX Sequence 1005 AA;

```

Query Match 9.2%; Score 248.5; DB 6; Length 1005;
 Best Local Similarity 21.6%; Pred. No. 5.3e-07;
 Matches 139; Conservative 117; Mismatches 197; Indels 191; Gaps 29;

```

QY 23 PAVTAE-----GPGSSQA-PRKEPGAQARTQAQSGALRDVSEELSRLE 64
Db 328 PAVTAEQEEBEOBQHVGTGTVNSVGSNCSIPSMISASSQSSVNSLPDVSDDKS-ELD 386
QY 65 DILSTYCVDNNOG-----GPGEDG--AQGEPAEPDAEKSRITYVAENGPEPTPVYGEKE 118
Db 387 XMEGDHTVMNSSVTHLKPEENYEEGDP-----RT---RASDQSPQVSRHKS 434
QY 119 PSKGDPNTEIRQSDVGDROHRRPOEKKKAKGLGKLT-----LLMOTLNTLTSTPEEKL 174
Db 435 HYRNRHFPATIRTSILV-----TRQOEHEQDSSELREQMSGYRMRROHQKQLMTLENK 490
QY 175 ALCKKYA-----ELLEHERN--SOKQMLLQKQSOVLQOE-----KDLRGESHK 217
Db 491 AEMDEHRLDKDLETOFQNNFAAEMKLIKXHQAAEKAQVMSNEEKFOHQHQAQKK 550
QY 218 AVLARKSLSLCRLQHRNRSLEKEGVQR-----AREEER--KRKEVTSHFQV--TLNDI 268
Db 551 EL--NSFLSSQKREYKLRKEQLKEELNENQSTPKKEKQEMLSKQKENTQHFQREBEANLL 608
QY 269 QLQME-----QHN-----ERNSKLRQENNELAERLKKLEQYELRE 304
Db 609 RQQRQVLELCRRFKRBMMLGRHNLQDILVREELNKRQTKQDLEHALLRQESMQELBF 668
QY 305 EHDVFKHQLQOOLVDAKLAQQAQEMLKBAERHOREKDFLLKEAVESQRCMLMKQOF 364
Db 669 RHLNTIQK---NRCELL--RLQFQTELTNQLEYNKRERERELSRKHVYMRQOPKSLKSKE 723
QY 365 THLKOOL-----ALYTEKPEEFQNTLSKSSEVPTTKQEMKMTKKIKLEKETMYRSRW 420
Db 724 LQIKQFQDTCKIQTYQYKALRNHLE-----TTPKSEHKAVLKRK--EETRKLAILA 776
QY 421 ESSNKALLMEAEKTVR-----DKELE 442
Db 777 EQYDHSINEMLSQALRLDEAQAECQVLKQQLQOELELLNAYQSKIKWCAQAQHDREL 836
QY 443 GLQVKI-----QLEKLCRALQTERND-----LNKRVODLSAGQGS-----L 480
Db 837 ELEQVSLRLAELEQKTEEMALQNERTERISLLERQARETEAFDSMRLGFSNMVL 896
QY 481 TDSGPERPBGQAQAPSSPRVTEAPCYGAP--STEASQOTGP 522
Db 897 SNLSPE-----AFSHSYFGASGWSHNPTGGPGP 924

RESULT 49
AAO30953
ID AAO30953 standard; protein; 1005 AA.
XX
XX AAO30953;
XX
XX 06-OCT-2003 (first entry)
XX
XX Human TAOJIX protein #3 used to modulate beta-catenin pathway.
XX
XX Human; beta-catenin; TAOJIX; gene therapy; endometrial cancer; breast;
XX bladder; stomach stromal sarcoma; cytostatic.
XX Homo sapiens.
XX
XX Key Location/Qualifiers
XX Domain 32..285
XX /note= "Kinase domain"
XX
XX W02003051905-A2.
XX
XX 26-JUN-2003.
XX
XX 12-DEC-2002; 2002WO-US039742.
XX
XX 13-DEC-2001; 2001US-0340312P.
XX

```


GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

QM protein - protein search, using sw model

Run on: June 7, 2004, 14:37:41 ; Search time 127 Seconds

(without alignments)
1174.091 Million cell updates/sec

Title: US-10-023-529-8

Perfect score: 2702

Sequence: 1 KSSPGQPEAGPEGAQERPSQ.....APSTEASGQTGPQETPSA 530

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1155919 seqs, 281338677 residues

Total number of hits satisfying chosen parameters: 1155919

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 150 summaries

Database : Published Applications AA*

1: /cgn2_6/prodata/1/pubaa/US07_PUBCOMB.pep.*
2: /cgn2_6/prodata/1/pubaa/PCT_NEW_PUB.pep.*
3: /cgn2_6/prodata/1/pubaa/US06_PUBCOMB.pep.*
4: /cgn2_6/prodata/1/pubaa/US06_PUBCOMB.pep.*
5: /cgn2_6/prodata/1/pubaa/US07_NEW_PUB.pep.*
6: /cgn2_6/prodata/1/pubaa/PCTUS_PUBCOMB.pep.*
7: /cgn2_6/prodata/1/pubaa/US08_NEW_PUB.pep.*
8: /cgn2_6/prodata/1/pubaa/US08_PUBCOMB.pep.*
9: /cgn2_6/prodata/1/pubaa/US09A_PUBCOMB.pep.*
10: /cgn2_6/prodata/1/pubaa/US09B_PUBCOMB.pep.*
11: /cgn2_6/prodata/1/pubaa/US09C_PUBCOMB.pep.*
12: /cgn2_6/prodata/1/pubaa/US09_NEW_PUB.pep.*
13: /cgn2_6/prodata/1/pubaa/US10A_PUBCOMB.pep.*
14: /cgn2_6/prodata/1/pubaa/US10B_PUBCOMB.pep.*
15: /cgn2_6/prodata/1/pubaa/US10C_PUBCOMB.pep.*
16: /cgn2_6/prodata/1/pubaa/US10_NEW_PUB.pep.*
17: /cgn2_6/prodata/1/pubaa/US60_NEW_PUB.pep.*
18: /cgn2_6/prodata/1/pubaa/US60_PUBCOMB.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	2702	100.0	530	9	US-09-962-055-8
2	2702	100.0	530	9	US-09-976-740-8
3	2702	100.0	530	12	US-10-671-242-8
4	2702	100.0	530	13	US-10-023-529-8
5	2702	100.0	530	13	US-10-023-523-8
6	2702	100.0	530	15	US-10-616-187-8
7	2693	99.7	546	9	US-09-976-740-44
8	2693	99.7	546	12	US-10-671-242-44
9	2693	99.7	546	13	US-10-023-529-44
10	2693	99.7	546	13	US-10-023-523-44
11	2693	99.7	546	15	US-10-616-187-44
12	2465.5	91.2	557	9	US-09-962-055-5
13	2465.5	91.2	557	9	US-09-976-740-5
14	2465.5	91.2	557	12	US-10-671-242-5
15	2465.5	91.2	557	13	US-10-023-529-5

16	2465.5	91.2	557	13	US-10-023-523-5	Sequence 5, Appli
17	2465.5	91.2	557	15	US-10-616-187-5	Sequence 5, Appli
18	234	86.0	510	12	US-10-256-774-2134	Sequence 2134, Ap
19	501	18.5	437	12	US-10-425-114-55188	Sequence 55188, A
20	468	17.3	436	12	US-10-424-599-207434	Sequence 207434,
21	444.5	16.5	314	12	US-10-425-114-43684	Sequence 43684, A
22	375	13.9	346	15	US-10-369-493-5130	Sequence 5130, Ap
23	304.5	11.3	206	12	US-10-425-114-45419	Sequence 45419, A
24	259	9.6	886	15	US-10-369-493-1016	Sequence 1016, Ap
25	257.5	9.5	860	12	US-10-072-012-838	Sequence 838, App
26	257.5	9.5	860	12	US-10-037-417-59	Sequence 59, Appli
27	257.5	9.5	860	12	US-10-080-334-166	Sequence 166, App
28	256	9.5	2871	14	US-10-146-473-41	Sequence 41, Appli
29	253.5	9.4	1881	14	US-10-032-585-7646	Sequence 7646, Ap
30	252.5	9.3	880	15	US-10-369-493-21643	Sequence 21643, A
31	248.5	9.2	786	10	US-09-291-417-23	Sequence 23, Appli
32	248.5	9.2	1001	10	US-09-291-417-31	Sequence 31, Appli
33	248.5	9.2	1005	14	US-10-177-233-254	Sequence 254, App
34	248.5	9.2	1005	14	US-10-317-835-18	Sequence 18, Appli
35	247	9.1	85	15	US-10-264-049-3413	Sequence 3413, Ap
36	247	9.1	1001	15	US-10-445-735-2	Sequence 2, Appli
37	244.5	9.0	909	9	US-09-925-299-988	Sequence 988, App
38	244.5	9.0	909	10	US-09-925-299-988	Sequence 1, Appli
39	242	9.0	1534	14	US-10-342-136-1	Sequence 102, App
40	238	8.8	892	14	US-10-205-219-102	Sequence 102, App
41	237	8.8	1938	16	US-10-408-765A-1168	Sequence 1168, Ap
42	236.5	8.8	1137	12	US-10-336-472-16	Sequence 16, Appli
43	236.5	8.8	3225	16	US-10-408-765A-254	Sequence 254, App
44	236	8.7	1938	16	US-10-408-765A-1629	Sequence 1629, Ap
45	235.5	8.7	1939	16	US-10-408-765A-2188	Sequence 2188, Ap
46	234.5	8.7	650	15	US-10-104-047-3636	Sequence 3636, Ap
47	234	8.7	1859	12	US-10-336-472-20	Sequence 20, Appli
48	234	8.7	1935	12	US-10-336-472-22	Sequence 22, Appli
49	233.5	8.6	512	15	US-10-108-260A-4080	Sequence 4080, Ap
50	230.5	8.5	3878	14	US-10-080-608A-11	Sequence 11, Appli
51	230.5	8.5	3899	14	US-10-171-311-4	Sequence 4, Appli
52	230.5	8.5	3907	14	US-10-171-311-2	Sequence 2, Appli
53	230.5	8.5	3811	15	US-10-370-685-100	Sequence 100, App
54	230.5	8.5	3911	16	US-10-408-765A-1839	Sequence 1839, Ap
55	230.5	8.5	3917	14	US-10-171-311-8	Sequence 8, Appli
56	230.5	8.5	3925	14	US-10-171-311-6	Sequence 6, Appli
57	228.5	8.5	1941	12	US-10-188-186-96	Sequence 96, Appli
58	228	8.4	1164	15	US-10-369-493-6564	Sequence 6564, Ap
59	228	8.4	2383	9	US-10-082-830-260	Sequence 260, App
60	227.5	8.4	660	9	US-09-864-761-47959	Sequence 47959, A
61	227.5	8.4	670	15	US-10-264-049-4653	Sequence 4653, Ap
62	227.5	8.4	828	15	US-10-029-386-32324	Sequence 32324, A
63	227.5	8.4	1038	14	US-10-032-585-7776	Sequence 7776, Ap
64	227.5	8.4	1162	12	US-09-894-273-2	Sequence 2, Appli
65	227.5	8.4	1162	14	US-10-294-804-2	Sequence 2, Appli
66	227	8.4	888	10	US-09-893-519A-73	Sequence 73, Appli
67	226.5	8.4	698	15	US-10-094-749-1328	Sequence 1928, Ap
68	226.5	8.4	1114	16	US-10-408-765A-2119	Sequence 2119, Ap
69	226.5	8.4	1790	15	US-10-363-493-1586	Sequence 1586, Ap
70	226.5	8.4	1940	10	US-09-738-630-99	Sequence 99, Appli
71	226	8.4	1999	15	US-10-028-248A-107	Sequence 107, App
72	226	8.4	1999	15	US-10-107-782-107	Sequence 107, App
73	224	8.3	1180	15	US-10-369-493-10206	Sequence 10206, A
74	224	8.3	1961	15	US-10-028-248A-105	Sequence 105, App
75	224	8.3	1961	15	US-10-107-782-105	Sequence 105, App
76	223.5	8.3	670	12	US-09-298-523B-63	Sequence 63, Appli
77	223.5	8.3	735	12	US-09-748-875-63	Sequence 63, Appli
78	223.5	8.3	735	12	US-10-087-192-831	Sequence 831, App
79	223.5	8.3	868	9	US-09-884-001-19	Sequence 19, Appli
80	223.5	8.3	1959	15	US-10-028-248A-36	Sequence 36, Appli
81	223.5	8.3	1959	15	US-10-107-782-36	Sequence 36, Appli
82	223	8.3	1379	14	US-10-205-219-5	Sequence 5, Appli
83	223	8.3	1974	15	US-10-369-493-6395	Sequence 6395, Ap
84	222.5	8.2	701	10	US-09-298-523B-62	Sequence 62, Appli
85	222.5	8.2	701	12	US-09-748-875-62	Sequence 62, Appli
86	222.5	8.2	1940	16	US-10-408-765A-1175	Sequence 1175, Ap
87	221.5	8.2	1043	15	US-10-310-154-449	Sequence 449, App
88	221	8.2	1091	15	US-10-369-493-6328	Sequence 6328, Ap

Db 361 KQOETHLKQALALYTEKPEFQNTLSKSSSEVFTTFKQEMKMTKKIKKLEKETTMYRSRW 420
Qy 421 ESSNKALLEMAEKTVRDKELEGLOVKIORLEKLCRALQTERNDLNKRVDLSAGGQSL 480
Db 421 ESSNKALLEMAEKTVRDKELEGLOVKIORLEKLCRALQTERNDLNKRVDLSAGGQSL 480
Qy 481 TDSGPERRPEGGAQAPSSPRVTEAPCYPCGAPSTASGOTGQPEPTSARA 530
Db 481 TDSGPERRPEGGAQAPSSPRVTEAPCYPCGAPSTASGOTGQPEPTSARA 530

RESULT 2
US-09-976-740-8
; Sequence 8, Application US/09976740
; Publication No. US20020194633A1
; GENERAL INFORMATION:
; APPLICANT: Lees, Ann M.
; APPLICANT: Lees, Robert S.
; APPLICANT: Law, Simon W.
; APPLICANT: Arjona, Anibal A.
; TITLE OF INVENTION: NOVEL LOW DENSITY LIPOPROTEIN BINDING
; TITLE OF INVENTION: PROTEINS AND THEIR USE IN DIAGNOSING AND TREATING
; TITLE OF INVENTION: ATHEROSCLEROSIS
; FILE REFERENCE: 10797-004001
; CURRENT APPLICATION NUMBER: US/09/976,740
; CURRENT FILING DATE: 2001-10-12
; PRIOR APPLICATION NUMBER: 09/616,289
; PRIOR FILING DATE: 2000-07-14
; PRIOR APPLICATION NUMBER: US 08/979,608
; PRIOR FILING DATE: 1997-11-26
; PRIOR APPLICATION NUMBER: US 60/031,930
; PRIOR FILING DATE: 1996-11-27
; PRIOR APPLICATION NUMBER: US 60/048,547
; PRIOR FILING DATE: 1997-06-03
; NUMBER OF SEQ ID NOS: 53
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 8
; LENGTH: 530
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-976-740-8

Query Match 100.0%; Score 2702; DB 9; Length 530;
Best Local Similarity 100.0%; Pred. No. 3.3e-156;
Matches 530; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 1 KSSPGQPEAGPEGAQERPSQAAPAVEAEGFGSSQAAPKPEGAQARTAQSGALRDVSELS 60
Db 1 KSSPGQPEAGPEGAQERPSQAAPAVEAEGFGSSQAAPKPEGAQARTAQSGALRDVSELS 60
Qy 61 RQLEDILSTYCVNNQGGPDGAGGEPAPEDAEKSRITYVARNGEPEPTPVVYGEKPS 120
Db 61 RQLEDILSTYCVNNQGGPDGAGGEPAPEDAEKSRITYVARNGEPEPTPVVYGEKPS 120
Qy 121 KGDPTTEIRQSDVGDGRDHRPQEKKAAGLGKEITLLMQTLNTLSTPEKLAALCKKY 180
Db 121 KGDPTTEIRQSDVGDGRDHRPQEKKAAGLGKEITLLMQTLNTLSTPEKLAALCKKY 180
Qy 181 AELLEHRNSQOMKLLQKQSQOLVQEKDHLRGHSHKAVLARSKLESCLRELQHRNSLK 240
Db 181 AELLEHRNSQOMKLLQKQSQOLVQEKDHLRGHSHKAVLARSKLESCLRELQHRNSLK 240
Qy 241 BEGVQARAREEKEKKEVTSHPQVTLNDIQLOMEQHNRNSKLRQENNELAERLKKLIEQY 300
Db 241 BEGVQARAREEKEKKEVTSHPQVTLNDIQLOMEQHNRNSKLRQENNELAERLKKLIEQY 300
Qy 301 ELREEHIDKVPKHDLOQQLVDKLAQQAQEMLKEABERHOREKDFLLKEAVESQRMCELM 360
Db 301 ELREEHIDKVPKHDLOQQLVDKLAQQAQEMLKEABERHOREKDFLLKEAVESQRMCELM 360
Qy 361 KQOETHLKQALALYTEKPEFQNTLSKSSSEVFTTFKQEMKMTKKIKKLEKETTMYRSRW 420
Db 361 KQOETHLKQALALYTEKPEFQNTLSKSSSEVFTTFKQEMKMTKKIKKLEKETTMYRSRW 420

Qy 421 ESSNKALLEMAEKTVRDKELEGLOVKIORLEKLCRALQTERNDLNKRVDLSAGGQSL 480
Db 421 ESSNKALLEMAEKTVRDKELEGLOVKIORLEKLCRALQTERNDLNKRVDLSAGGQSL 480
Qy 481 TDSGPERRPEGGAQAPSSPRVTEAPCYPCGAPSTASGOTGQPEPTSARA 530
Db 481 TDSGPERRPEGGAQAPSSPRVTEAPCYPCGAPSTASGOTGQPEPTSARA 530

RESULT 3
US-10-671-242-8
; Sequence 8, Application US/10671242
; Publication No. US20040040049A1
; GENERAL INFORMATION:
; APPLICANT: Lees, Ann M.
; APPLICANT: Lees, Robert S.
; APPLICANT: Law, Simon W.
; APPLICANT: Arjona, Anibal A.
; TITLE OF INVENTION: NOVEL LOW DENSITY LIPOPROTEIN BINDING
; TITLE OF INVENTION: PROTEINS AND THEIR USE IN DIAGNOSING AND TREATING
; TITLE OF INVENTION: ATHEROSCLEROSIS
; FILE REFERENCE: 10797-004001
; CURRENT APPLICATION NUMBER: US/10/671,242
; CURRENT FILING DATE: 2003-09-24
; PRIOR APPLICATION NUMBER: US/09/616,289
; PRIOR FILING DATE: 2000-07-14
; PRIOR APPLICATION NUMBER: US 09/517,849
; PRIOR FILING DATE: 2000-03-02
; PRIOR APPLICATION NUMBER: US 08/979,608
; PRIOR FILING DATE: 1997-11-26
; PRIOR APPLICATION NUMBER: US 60/031,930
; PRIOR FILING DATE: 1996-11-27
; PRIOR APPLICATION NUMBER: US 60/048,547
; PRIOR FILING DATE: 1997-06-03
; NUMBER OF SEQ ID NOS: 53
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 8
; LENGTH: 530
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-671-242-8

Query Match 100.0%; Score 2702; DB 12; Length 530;
Best Local Similarity 100.0%; Pred. No. 3.3e-156;
Matches 530; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 1 KSSPGQPEAGPEGAQERPSQAAPAVEAEGFGSSQAAPKPEGAQARTAQSGALRDVSELS 60
Db 1 KSSPGQPEAGPEGAQERPSQAAPAVEAEGFGSSQAAPKPEGAQARTAQSGALRDVSELS 60
Qy 61 RQLEDILSTYCVNNQGGPDGAGGEPAPEDAEKSRITYVARNGEPEPTPVVYGEKPS 120
Db 61 RQLEDILSTYCVNNQGGPDGAGGEPAPEDAEKSRITYVARNGEPEPTPVVYGEKPS 120
Qy 121 KGDPTTEIRQSDVGDGRDHRPQEKKAAGLGKEITLLMQTLNTLSTPEKLAALCKKY 180
Db 121 KGDPTTEIRQSDVGDGRDHRPQEKKAAGLGKEITLLMQTLNTLSTPEKLAALCKKY 180
Qy 181 AELLEHRNSQOMKLLQKQSQOLVQEKDHLRGHSHKAVLARSKLESCLRELQHRNSLK 240
Db 181 AELLEHRNSQOMKLLQKQSQOLVQEKDHLRGHSHKAVLARSKLESCLRELQHRNSLK 240
Qy 241 BEGVQARAREEKEKKEVTSHPQVTLNDIQLOMEQHNRNSKLRQENNELAERLKKLIEQY 300
Db 241 BEGVQARAREEKEKKEVTSHPQVTLNDIQLOMEQHNRNSKLRQENNELAERLKKLIEQY 300
Qy 301 ELREEHIDKVPKHDLOQQLVDKLAQQAQEMLKEABERHOREKDFLLKEAVESQRMCELM 360
Db 301 ELREEHIDKVPKHDLOQQLVDKLAQQAQEMLKEABERHOREKDFLLKEAVESQRMCELM 360
Qy 361 KQOETHLKQALALYTEKPEFQNTLSKSSSEVFTTFKQEMKMTKKIKKLEKETTMYRSRW 420
Db 361 KQOETHLKQALALYTEKPEFQNTLSKSSSEVFTTFKQEMKMTKKIKKLEKETTMYRSRW 420

Db 361 KQETHLKKQALALYTERKFEFQNTLSKSSSEVFTTFKQEMKMTKKIKLEKETTTWYRSW 420
Qy 421 ESSNKALLEMAEKEKTVRDKEGLGVKIQRIKLEKCRALQTERNDLNKRVODLSAGGQSL 480
Db 421 ESSNKALLEMAEKEKTVRDKEGLGVKIQRIKLEKCRALQTERNDLNKRVODLSAGGQSL 480
Qy 481 TDSGPRRPEGPGCAQAPSSPRVTAPCPYGPAPSTASGQTGPQEPPTSARA 530
Db 481 TDSGPRRPEGPGCAQAPSSPRVTAPCPYGPAPSTASGQTGPQEPPTSARA 530

RESULT 4

US-10-023-529-8
; Sequence 8, Application US/10023529
; Publication No. US20020129388A1
; GENERAL INFORMATION:
; APPLICANT: Lees, Ann M.
; APPLICANT: Lees, Robert S.
; APPLICANT: Law, Simon W.
; APPLICANT: Arjona, Anibal A.
; TITLE OF INVENTION: NOVEL LOW DENSITY LIPOPROTEIN BINDING
; TITLE OF INVENTION: PROTEINS AND THEIR USE IN DIAGNOSING AND TREATING
; TITLE OF INVENTION: ATHEROSCLEROSIS
; FILE REFERENCE: 10797-004001
; CURRENT APPLICATION NUMBER: US/10/023,529
; CURRENT FILING DATE: 2001-12-17
; PRIOR APPLICATION NUMBER: 09/616,289
; PRIOR FILING DATE: 2000-07-14
; PRIOR APPLICATION NUMBER: US 09/517,849
; PRIOR FILING DATE: 2000-03-02
; PRIOR APPLICATION NUMBER: US 08/979,608
; PRIOR FILING DATE: 1997-11-26
; PRIOR APPLICATION NUMBER: US 60/031,930
; PRIOR FILING DATE: 1996-11-27
; PRIOR APPLICATION NUMBER: US 60/048,547
; PRIOR FILING DATE: 1997-06-03
; NUMBER OF SEQ ID NOS: 53
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 8
; LENGTH: 530
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-023-529-8

Query Match 100.0%; Score 2702; DB 13; Length 530;
Best Local Similarity 100.0%; Pred. No. 3.3e-156;
Matches 530; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 1 KSSPGQPEAGPEGAQERPSQAAPAVEAGPGSSQAAPRKEGAQARTAQSGALRDVSELS 60
Db 1 KSSPGQPEAGPEGAQERPSQAAPAVEAGPGSSQAAPRKEGAQARTAQSGALRDVSELS 60
Qy 61 RQLEDILSTYCVDNNOGGPGDGAQGEPAEPEDAESKRTYVARNGEPEPTPVVYGEKPS 120
Db 61 RQLEDILSTYCVDNNOGGPGDGAQGEPAEPEDAESKRTYVARNGEPEPTPVVYGEKPS 120
Qy 121 KGDPTNTEIROSDEVDGDRHRRPQEKKAAGLKGKEITLLMOTLNTLSTPEKLAALCKY 180
Db 121 KGDPTNTEIROSDEVDGDRHRRPQEKKAAGLKGKEITLLMOTLNTLSTPEKLAALCKY 180
Qy 181 AEILLEHRSOKMKLLQKQSQVLQVQKDHRLRGEHSKAVLARSKLESLCRELQHRNLSLX 240
Db 181 AEILLEHRSOKMKLLQKQSQVLQVQKDHRLRGEHSKAVLARSKLESLCRELQHRNLSLX 240
Qy 241 EGVQARAEERKEKVTSHFQVTLNDIQLQMEQHNRNSKLQENNELAERLKKLIEQY 300
Db 241 EGVQARAEERKEKVTSHFQVTLNDIQLQMEQHNRNSKLQENNELAERLKKLIEQY 300
Qy 301 ELREHHIDKVPKHDLQOOLVDKLOQAQEMLKEAERHOREKDFLLKEAVESQRMCELM 360
Db 301 ELREHHIDKVPKHDLQOOLVDKLOQAQEMLKEAERHOREKDFLLKEAVESQRMCELM 360
Qy 361 KQETHLKKQALALYTERKFEFQNTLSKSSSEVFTTFKQEMKMTKKIKLEKETTTWYRSW 420

Db 361 KQETHLKKQALALYTERKFEFQNTLSKSSSEVFTTFKQEMKMTKKIKLEKETTTWYRSW 420
Qy 421 ESSNKALLEMAEKEKTVRDKEGLGVKIQRIKLEKCRALQTERNDLNKRVODLSAGGQSL 480
Db 421 ESSNKALLEMAEKEKTVRDKEGLGVKIQRIKLEKCRALQTERNDLNKRVODLSAGGQSL 480
Qy 481 TDSGPRRPEGPGCAQAPSSPRVTAPCPYGPAPSTASGQTGPQEPPTSARA 530
Db 481 TDSGPRRPEGPGCAQAPSSPRVTAPCPYGPAPSTASGQTGPQEPPTSARA 530

RESULT 5

US-10-023-523-8
; Sequence 8, Application US/10023523
; Publication No. US20020152485A1
; GENERAL INFORMATION:
; APPLICANT: Lees, Ann M.
; APPLICANT: Lees, Robert S.
; APPLICANT: Law, Simon W.
; APPLICANT: Arjona, Anibal A.
; TITLE OF INVENTION: NOVEL LOW DENSITY LIPOPROTEIN BINDING
; TITLE OF INVENTION: PROTEINS AND THEIR USE IN DIAGNOSING AND TREATING
; TITLE OF INVENTION: ATHEROSCLEROSIS
; FILE REFERENCE: 10797-004001
; CURRENT APPLICATION NUMBER: US/10/023,523
; CURRENT FILING DATE: 2001-12-17
; PRIOR APPLICATION NUMBER: US/09/616,289
; PRIOR FILING DATE: 2000-07-14
; PRIOR APPLICATION NUMBER: US 09/517,849
; PRIOR FILING DATE: 2000-03-02
; PRIOR APPLICATION NUMBER: US 08/979,608
; PRIOR FILING DATE: 1997-11-26
; PRIOR APPLICATION NUMBER: US 60/031,930
; PRIOR FILING DATE: 1996-11-27
; PRIOR APPLICATION NUMBER: US 60/048,547
; PRIOR FILING DATE: 1997-06-03
; NUMBER OF SEQ ID NOS: 53
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 8
; LENGTH: 530
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-023-523-8

Query Match 100.0%; Score 2702; DB 13; Length 530;
Best Local Similarity 100.0%; Pred. No. 3.3e-156;
Matches 530; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 1 KSSPGQPEAGPEGAQERPSQAAPAVEAGPGSSQAAPRKEGAQARTAQSGALRDVSELS 60
Db 1 KSSPGQPEAGPEGAQERPSQAAPAVEAGPGSSQAAPRKEGAQARTAQSGALRDVSELS 60
Qy 61 RQLEDILSTYCVDNNOGGPGDGAQGEPAEPEDAESKRTYVARNGEPEPTPVVYGEKPS 120
Db 61 RQLEDILSTYCVDNNOGGPGDGAQGEPAEPEDAESKRTYVARNGEPEPTPVVYGEKPS 120
Qy 121 KGDPTNTEIROSDEVDGDRHRRPQEKKAAGLKGKEITLLMOTLNTLSTPEKLAALCKY 180
Db 121 KGDPTNTEIROSDEVDGDRHRRPQEKKAAGLKGKEITLLMOTLNTLSTPEKLAALCKY 180
Qy 181 AEILLEHRSOKMKLLQKQSQVLQVQKDHRLRGEHSKAVLARSKLESLCRELQHRNLSLX 240
Db 181 AEILLEHRSOKMKLLQKQSQVLQVQKDHRLRGEHSKAVLARSKLESLCRELQHRNLSLX 240
Qy 241 EGVQARAEERKEKVTSHFQVTLNDIQLQMEQHNRNSKLQENNELAERLKKLIEQY 300
Db 241 EGVQARAEERKEKVTSHFQVTLNDIQLQMEQHNRNSKLQENNELAERLKKLIEQY 300
Qy 301 ELREHHIDKVPKHDLQOOLVDKLOQAQEMLKEAERHOREKDFLLKEAVESQRMCELM 360
Db 301 ELREHHIDKVPKHDLQOOLVDKLOQAQEMLKEAERHOREKDFLLKEAVESQRMCELM 360

QY 361 KQETHLKQALALYTEKPEFQNTLSKSSSEVFTTFKQEMKMTKKIKKLEKETTYHSRW 420
DB 361 KQETHLKQALALYTEKPEFQNTLSKSSSEVFTTFKQEMKMTKKIKKLEKETTYHSRW 420
QY 421 ESSNKALLEMAEKTVDKLEGLQVKIORLEKLCRALQTERNDLNKRVQDLSAGGGS 480
DB 421 ESSNKALLEMAEKTVDKLEGLQVKIORLEKLCRALQTERNDLNKRVQDLSAGGGS 480
QY 481 TDSGPERRPEGPGQAQAPSSPRVTEAPCYPCAPSTEASGQTGPQPTTSARA 530
DB 481 TDSGPERRPEGPGQAQAPSSPRVTEAPCYPCAPSTEASGQTGPQPTTSARA 530

RESULT 6

US-10-616-187-8
; Sequence 8, Application US/10616187
; Publication No. US20040013668A1
; GENERAL INFORMATION:
; APPLICANT: Lees, Ann M.
; APPLICANT: Lees, Robert S.
; APPLICANT: Law, Simon W.
; APPLICANT: Arjona, Anibal A.
; TITLE OF INVENTION: NOVEL LOW DENSITY LIPOPROTEIN BINDING
; TITLE OF INVENTION: PROTEINS AND THEIR USE IN DIAGNOSING AND TREATING
; TITLE OF INVENTION: ATHEROSCLEROSIS
; FILE REFERENCE: 10797-004001
; CURRENT APPLICATION NUMBER: US/10/616,187
; CURRENT FILING DATE: 2003-07-09
; PRIOR APPLICATION NUMBER: US/09/616,289
; PRIOR FILING DATE: 2000-07-14
; PRIOR APPLICATION NUMBER: US 09/517,849
; PRIOR FILING DATE: 2000-03-02
; PRIOR APPLICATION NUMBER: US 08/979,608
; PRIOR FILING DATE: 1997-11-26
; PRIOR APPLICATION NUMBER: US 60/031,930
; PRIOR FILING DATE: 1996-11-27
; PRIOR APPLICATION NUMBER: US 60/048,547
; PRIOR FILING DATE: 1997-06-03
; NUMBER OF SEQ ID NOS: 53
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 8
; LENGTH: 530
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-616-187-8

Query Match 100.0%; Score 2702; DB 15; Length 530;
Best Local Similarity 100.0%; Pred. No. 3.3e-156;
Matches 530; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 KSSPGQPEAGPEGAQRRPSQAAPAVEAGPGSSQAAPRKPGEQAQRTAQSGALRDVSEELS 60
DB 1 KSSPGQPEAGPEGAQRRPSQAAPAVEAGPGSSQAAPRKPGEQAQRTAQSGALRDVSEELS 60
QY 61 RQLEDILSTYCVNNQGGPGEDGAQGEPAEPEDAERKRTYVARNGEPEPTPVVYGEKPS 120
DB 61 RQLEDILSTYCVNNQGGPGEDGAQGEPAEPEDAERKRTYVARNGEPEPTPVVYGEKPS 120
QY 121 KGDPTNTEIRQSDVGDHRDRHPQKKAAGLKGKLTLLMTQTLNTLSTPEKLAALCKKY 180
DB 121 KGDPTNTEIRQSDVGDHRDRHPQKKAAGLKGKLTLLMTQTLNTLSTPEKLAALCKKY 180
QY 181 AELLFEHRNSQKMKLLQKQSQVLVQEKDHLRGHSHKAVLARSKLBSLCRELQHRNSLK 240
DB 181 AELLFEHRNSQKMKLLQKQSQVLVQEKDHLRGHSHKAVLARSKLBSLCRELQHRNSLK 240
QY 241 EGVQVAREEERKEKVTSHFQVTLNDIQLQMEQHNRNSKLRQENNELAERLKLIEQY 300
DB 241 EGVQVAREEERKEKVTSHFQVTLNDIQLQMEQHNRNSKLRQENNELAERLKLIEQY 300
QY 301 ELREEHIDKVPFKHKLQOQLVDAKLQQAQEMLKEAERHOREKDFLLKEAVESQRMCELM 360
DB 301 ELREEHIDKVPFKHKLQOQLVDAKLQQAQEMLKEAERHOREKDFLLKEAVESQRMCELM 360

QY 361 KQETHLKQALALYTEKPEFQNTLSKSSSEVFTTFKQEMKMTKKIKKLEKETTYHSRW 420
DB 361 KQETHLKQALALYTEKPEFQNTLSKSSSEVFTTFKQEMKMTKKIKKLEKETTYHSRW 420
QY 421 ESSNKALLEMAEKTVDKLEGLQVKIORLEKLCRALQTERNDLNKRVQDLSAGGGS 480
DB 421 ESSNKALLEMAEKTVDKLEGLQVKIORLEKLCRALQTERNDLNKRVQDLSAGGGS 480
QY 481 TDSGPERRPEGPGQAQAPSSPRVTEAPCYPCAPSTEASGQTGPQPTTSARA 530
DB 481 TDSGPERRPEGPGQAQAPSSPRVTEAPCYPCAPSTEASGQTGPQPTTSARA 530

RESULT 7

US-09-976-740-44
; Sequence 44, Application US/09976740
; Publication No. US20020194633A1
; GENERAL INFORMATION:
; APPLICANT: Lees, Ann M.
; APPLICANT: Lees, Robert S.
; APPLICANT: Law, Simon W.
; APPLICANT: Arjona, Anibal A.
; TITLE OF INVENTION: NOVEL LOW DENSITY LIPOPROTEIN BINDING
; TITLE OF INVENTION: PROTEINS AND THEIR USE IN DIAGNOSING AND TREATING
; TITLE OF INVENTION: ATHEROSCLEROSIS
; FILE REFERENCE: 10797-004001
; CURRENT APPLICATION NUMBER: US/09/976,740
; CURRENT FILING DATE: 2001-10-12
; PRIOR APPLICATION NUMBER: 09/616,289
; PRIOR FILING DATE: 2000-07-14
; PRIOR APPLICATION NUMBER: US 08/979,608
; PRIOR FILING DATE: 1997-11-26
; PRIOR APPLICATION NUMBER: US 60/031,930
; PRIOR FILING DATE: 1996-11-27
; PRIOR APPLICATION NUMBER: US 60/048,547
; PRIOR FILING DATE: 1997-06-03
; NUMBER OF SEQ ID NOS: 53
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 44
; LENGTH: 546
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-976-740-44

Query Match 99.7%; Score 2693; DB 9; Length 546;
Best Local Similarity 99.8%; Pred. No. 1.2e-155;
Matches 529; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 1 KSSPGQPEAGPEGAQRRPSQAAPAVEAGPGSSQAAPRKPGEQAQRTAQSGALRDVSEELS 60
DB 17 KSSPGQPEAGPEGAQRRPSQAAPAVEAGPGSSQAAPRKPGEQAQRTAQSGALRDVSEELS 76
QY 61 RQLEDILSTYCVNNQGGPGEDGAQGEPAEPEDAERKRTYVARNGEPEPTPVVYGEKPS 120
DB 77 RQLEDILSTYCVNNQGGPGEDGAQGEPAEPEDAERKRTYVARNGEPEPTPVVYGEKPS 136
QY 121 KGDPTNTEIRQSDVGDHRDRHPQKKAAGLKGKLTLLMTQTLNTLSTPEKLAALCKKY 180
DB 137 KGDPTNTEIRQSDVGDHRDRHPQKKAAGLKGKLTLLMTQTLNTLSTPEKLAALCKKY 196
QY 181 AELLFEHRNSQKMKLLQKQSQVLVQEKDHLRGHSHKAVLARSKLBSLCRELQHRNSLK 240
DB 197 AELLFEHRNSQKMKLLQKQSQVLVQEKDHLRGHSHKAVLARSKLBSLCRELQHRNSLK 256
QY 241 EGVQVAREEERKEKVTSHFQVTLNDIQLQMEQHNRNSKLRQENNELAERLKLIEQY 300
DB 257 EGVQVAREEERKEKVTSHFQVTLNDIQLQMEQHNRNSKLRQENNELAERLKLIEQY 316
QY 301 ELREEHIDKVPFKHKLQOQLVDAKLQQAQEMLKEAERHOREKDFLLKEAVESQRMCELM 360
DB 317 ELREEHIDKVPFKHKLQOQLVDAKLQQAQEMLKEAERHOREKDFLLKEAVESQRMCELM 376

Qy 361 KQETHLKQALALYTEKFEFQNTLSKSSSEVFTTFKQEMKMTKKIKLEKETTMTSRW 420
Db 377 KQETHLKQALALYTEKFEFQNTLSKSSSEVFTTFKQEMKMTKKIKLEKETTMTSRW 436
Qy 421 ESSNKALLEWAEKTVRDKEGLQVKIORLEKLCRALQTERNDLNKRVODLSAGGQSSL 480
Db 437 ESSNKALLEWAEKTVRDKEGLQVKIORLEKLCRALQTERNDLNKRVODLSAGGQSSL 496
Qy 481 TDSGPERRPEGGAQAPSSPRVTEAPCYPGAPSTEASGQTGPQPTTSARA 530
Db 497 TDSGPERRPEGGAQAPSSPRVTEAPCYPGAPSTEASGQTGPQPTTSARA 546

RESULT 8

US-10-671-242-44
; Sequence 44, Application US/10671242
; Publication No. US20040040049A1
; GENERAL INFORMATION:
; APPLICANT: Lees, Ann M.
; APPLICANT: Lees, Robert S.
; APPLICANT: Law, Simon W.
; APPLICANT: Arjona, Anibal A.
; TITLE OF INVENTION: NOVEL LOW DENSITY LIPOPROTEIN BINDING
; TITLE OF INVENTION: PROTEINS AND THEIR USE IN DIAGNOSING AND TREATING
; TITLE OF INVENTION: ATHEROSCLEROSIS
; FILE REFERENCE: 10797-004001
; CURRENT APPLICATION NUMBER: US/10/671,242
; CURRENT FILING DATE: 2003-09-24
; PRIOR APPLICATION NUMBER: US/09/616,289
; PRIOR FILING DATE: 2000-07-14
; PRIOR APPLICATION NUMBER: US 09/517,849
; PRIOR FILING DATE: 2000-03-02
; PRIOR APPLICATION NUMBER: US 08/979,608
; PRIOR FILING DATE: 1997-11-26
; PRIOR APPLICATION NUMBER: US 60/031,930
; PRIOR FILING DATE: 1996-11-27
; PRIOR APPLICATION NUMBER: US 60/048,547
; PRIOR FILING DATE: 1997-06-03
; NUMBER OF SEQ ID NOS: 53
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 44
; LENGTH: 546
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-671-242-44

Query Match 99.7%; Score 2693; DB 12; Length 546;
Best Local Similarity 99.8%; Pred. No. 1.2e-155;
Matches 529; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
Qy 1 KSSPGQPEAGPEGAQERPSQAAPAVEAEGPGSSQAAPRKGQAQARTAQSGALRDVSEELS 60
Db 17 KSSPGQPEAGPEGAQERPSQAAPAVEAEGPGSSQAAPRKGQAQARTAQSGALRDVSEELS 76
Qy 51 RQLEDILSTYCVDNNGGPGEDGAGGPAEPEDAERKSTYVARGGEPTPVVYGEKPS 120
Db 77 RQLEDILSTYCVDNNGGPGEDGAGGPAEPEDAERKSTYVARGGEPTPVVYGEKPS 136
Qy 121 KGDPTNTEIRQSDVEGDRDHRPQEKKAKGLGKEITLLMQTLNTLSTPEKLAALCKKY 180
Db 137 KGDPTNTEIRQSDVEGDRDHRPQEKKAKGLGKEITLLMQTLNTLSTPEKLAALCKKY 196
Qy 181 AELLIEHRNSQOMKLLQKQSQLVQKDHARGHSAVLARSKLESICRELQHRNSLK 240
Db 197 AELLIEHRNSQOMKLLQKQSQLVQKDHARGHSAVLARSKLESICRELQHRNSLK 256
Qy 241 BEGVQARAEEREEKKEVTSHFQVTLNDIQLQMEQHNRNSKLRQENNELAERLAKKLEQY 300
Db 257 BEGVQARAEEREEKKEVTSHFQVTLNDIQLQMEQHNRNSKLRQENNELAERLAKKLEQY 316
Qy 301 ELREBHIDKVPFKHDLQOOLVDKALQQAQEMLKEAERHOREKDFLLKBAVESQRCMLM 360
Db 317 ELREBHIDKVPFKHDLQOOLVDKALQQAQEMLKEAERHOREKDFLLKBAVESQRCMLM 376

Qy 361 KQETHLKQALALYTEKFEFQNTLSKSSSEVFTTFKQEMKMTKKIKLEKETTMTSRW 420
Db 377 KQETHLKQALALYTEKFEFQNTLSKSSSEVFTTFKQEMKMTKKIKLEKETTMTSRW 436
Qy 421 ESSNKALLEWAEKTVRDKEGLQVKIORLEKLCRALQTERNDLNKRVODLSAGGQSSL 480
Db 437 ESSNKALLEWAEKTVRDKEGLQVKIORLEKLCRALQTERNDLNKRVODLSAGGQSSL 496
Qy 481 TDSGPERRPEGGAQAPSSPRVTEAPCYPGAPSTEASGQTGPQPTTSARA 530
Db 497 TDSGPERRPEGGAQAPSSPRVTEAPCYPGAPSTEASGQTGPQPTTSARA 546

RESULT 9

US-10-023-529-44
; Sequence 44, Application US/10023529
; Publication No. US20020129388A1
; GENERAL INFORMATION:
; APPLICANT: Lees, Ann M.
; APPLICANT: Lees, Robert S.
; APPLICANT: Law, Simon W.
; APPLICANT: Arjona, Anibal A.
; TITLE OF INVENTION: NOVEL LOW DENSITY LIPOPROTEIN BINDING
; TITLE OF INVENTION: PROTEINS AND THEIR USE IN DIAGNOSING AND TREATING
; TITLE OF INVENTION: ATHEROSCLEROSIS
; FILE REFERENCE: 10797-004001
; CURRENT APPLICATION NUMBER: US/10/023,529
; CURRENT FILING DATE: 2001-12-17
; PRIOR APPLICATION NUMBER: 09/616,289
; PRIOR FILING DATE: 2000-07-14
; PRIOR APPLICATION NUMBER: US 09/517,849
; PRIOR FILING DATE: 2000-03-02
; PRIOR APPLICATION NUMBER: US 08/979,608
; PRIOR FILING DATE: 1997-11-26
; PRIOR APPLICATION NUMBER: US 60/031,930
; PRIOR FILING DATE: 1996-11-27
; PRIOR APPLICATION NUMBER: US 60/048,547
; PRIOR FILING DATE: 1997-06-03
; NUMBER OF SEQ ID NOS: 53
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 44
; LENGTH: 546
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-023-529-44

Query Match 99.7%; Score 2693; DB 13; Length 546;
Best Local Similarity 99.8%; Pred. No. 1.2e-155;
Matches 529; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
Qy 1 KSSPGQPEAGPEGAQERPSQAAPAVEAEGPGSSQAAPRKGQAQARTAQSGALRDVSEELS 60
Db 17 KSSPGQPEAGPEGAQERPSQAAPAVEAEGPGSSQAAPRKGQAQARTAQSGALRDVSEELS 76
Qy 61 RQLEDILSTYCVDNNGGPGEDGAGGPAEPEDAERKSTYVARGGEPTPVVYGEKPS 120
Db 77 RQLEDILSTYCVDNNGGPGEDGAGGPAEPEDAERKSTYVARGGEPTPVVYGEKPS 136
Qy 121 KGDPTNTEIRQSDVEGDRDHRPQEKKAKGLGKEITLLMQTLNTLSTPEKLAALCKKY 180
Db 137 KGDPTNTEIRQSDVEGDRDHRPQEKKAKGLGKEITLLMQTLNTLSTPEKLAALCKKY 196
Qy 181 AELLIEHRNSQOMKLLQKQSQLVQKDHARGHSAVLARSKLESICRELQHRNSLK 240
Db 197 AELLIEHRNSQOMKLLQKQSQLVQKDHARGHSAVLARSKLESICRELQHRNSLK 256
Qy 241 BEGVQARAEEREEKKEVTSHFQVTLNDIQLQMEQHNRNSKLRQENNELAERLAKKLEQY 300
Db 257 BEGVQARAEEREEKKEVTSHFQVTLNDIQLQMEQHNRNSKLRQENNELAERLAKKLEQY 316
Qy 301 ELREBHIDKVPFKHDLQOOLVDKALQQAQEMLKEAERHOREKDFLLKBAVESQRCMLM 360

Db 317 ELREHIDKVFKKHDLQQLVDKIQQAQEMLKEABERHOREKDFLLKEAVESQRMCELM 376
QY 361 KOQETHLKOQLALYTEKPEPONTLSKSEVFTTFKQEMERMTKKIKLEKETMTYRSRW 420
Db 377 KOQETHLKOQLALYTEKPEPONTLSKSEVFTTFKQEMERMTKKIKLEKETMTYRSRW 436
QY 421 ESSNKALLEMAEKKTVRDKLEGLQVKIQRLEKLCRALQTERNDLNKRVQDLISAGGQSSL 480
Db 437 ESSNKALLEMAEKKTVRDKLEGLQVKIQRLEKLCRALQTERNDLNKRVQDLISAGGQSSL 496
QY 481 TDGSPRRPPEGGAQAPSSPRVTEAPCYPGAPSTEASGQTGPQEPSARA 530
Db 497 TDGSPRRPPEGGAQAPSSPRVTEAPCYPGAPSTEASGQTGPQEPSARA 546
RESULT 10
US-10-023-523-44
; Sequence 44, Application US/10023523
; Publication No. US20020152485A1
; GENERAL INFORMATION:
; APPLICANT: Lees, Ann M.
; APPLICANT: Lees, Robert S.
; APPLICANT: Law, Simon W.
; APPLICANT: Arjona, Anibal A.
; TITLE OF INVENTION: NOVEL LOW DENSITY LIPOPROTEIN BINDING
; TITLE OF INVENTION: PROTEINS AND THEIR USE IN DIAGNOSING AND TREATING
; FILE REFERENCE: 10797-004001
; CURRENT APPLICATION NUMBER: US/10/023,523
; CURRENT FILING DATE: 2001-12-17
; PRIOR APPLICATION NUMBER: US/09/616,289
; PRIOR FILING DATE: 2000-07-14
; PRIOR FILING DATE: 2000-03-02
; PRIOR FILING DATE: 1997-11-26
; PRIOR FILING DATE: 1996-11-27
; PRIOR APPLICATION NUMBER: US 60/031,930
; PRIOR FILING DATE: 1997-06-03
; NUMBER OF SEQ ID NOS: 53
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 44
; LENGTH: 546
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-023-523-44

Query Match 99.7%; Score 2693; DB 13; Length 546;
Best Local Similarity 99.8%; Pred. No. 1.2e-155;
Matches 529; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 1 KSSPGQPEAGPEGAQERPSQAAPVAEAGPGSSQAAPRKEPGAQARTAQSGALRDVSEELS 60
Db 17 KSSPGQPEAGPEGAQERPSQAAPVAEAGPGSSQAAPRKEPGAQARTAQSGALRDVSEELS 76
QY 61 RQLEDILSTYCVDNQGGPDEGAQEPAPEDAEKSRITYVARGPEPTPVVYGEKPS 120
Db 77 RQLEDILSTYCVDNQGGPDEGAQEPAPEDAEKSRITYVARGPEPTPVVYGEKPS 136
QY 121 KGPNTTEIRQSDVGDHRRPQEKKAQKGLGKEITLLMQTLNTLSTPEEKLAALCKKY 180
Db 137 KGPNTTEIRQSDVGDHRRPQEKKAQKGLGKEITLLMQTLNTLSTPEEKLAALCKKY 196
QY 181 ABLEEHNSQOMKLLQKQSQOLVQKDHRLRGEHSKAVLARSKLSLCRELQRNRSIK 240
Db 197 ABLEEHNSQOMKLLQKQSQOLVQKDHRLRGEHSKAVLARSKLSLCRELQRNRSIK 256
QY 241 EGVQARAEERKEKVTSHFQVTLNDIQLQMEQHNRNSKLQRQENMELARLKKLIEQY 300
Db 257 EGVQARAEERKEKVTSHFQVTLNDIQLQMEQHNRNSKLQRQENMELARLKKLIEQY 316
QY 301 ELREHIDKVFKKHDLQQLVDKIQQAQEMLKEABERHOREKDFLLKEAVESQRMCELM 360

Db 317 ELREHIDKVFKKHDLQQLVDKIQQAQEMLKEABERHOREKDFLLKEAVESQRMCELM 376
QY 361 KOQETHLKOQLALYTEKPEPONTLSKSEVFTTFKQEMERMTKKIKLEKETMTYRSRW 420
Db 377 KOQETHLKOQLALYTEKPEPONTLSKSEVFTTFKQEMERMTKKIKLEKETMTYRSRW 436
QY 421 ESSNKALLEMAEKKTVRDKLEGLQVKIQRLEKLCRALQTERNDLNKRVQDLISAGGQSSL 480
Db 437 ESSNKALLEMAEKKTVRDKLEGLQVKIQRLEKLCRALQTERNDLNKRVQDLISAGGQSSL 496
QY 481 TDGSPRRPPEGGAQAPSSPRVTEAPCYPGAPSTEASGQTGPQEPSARA 530
Db 497 TDGSPRRPPEGGAQAPSSPRVTEAPCYPGAPSTEASGQTGPQEPSARA 546
RESULT 11
US-10-616-187-44
; Sequence 44, Application US/10616187
; Publication No. US20040013668A1
; GENERAL INFORMATION:
; APPLICANT: Lees, Ann M.
; APPLICANT: Lees, Robert S.
; APPLICANT: Law, Simon W.
; APPLICANT: Arjona, Anibal A.
; TITLE OF INVENTION: NOVEL LOW DENSITY LIPOPROTEIN BINDING
; TITLE OF INVENTION: PROTEINS AND THEIR USE IN DIAGNOSING AND TREATING
; FILE REFERENCE: 10797-004001
; CURRENT APPLICATION NUMBER: US/10/616,187
; CURRENT FILING DATE: 2003-07-09
; PRIOR APPLICATION NUMBER: US/09/616,289
; PRIOR FILING DATE: 2000-07-14
; PRIOR APPLICATION NUMBER: US 09/517,849
; PRIOR FILING DATE: 2000-03-02
; PRIOR APPLICATION NUMBER: US 08/979,608
; PRIOR FILING DATE: 1997-11-26
; PRIOR APPLICATION NUMBER: US 60/031,930
; PRIOR FILING DATE: 1996-11-27
; PRIOR APPLICATION NUMBER: US 60/048,547
; PRIOR FILING DATE: 1997-06-03
; NUMBER OF SEQ ID NOS: 53
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 44
; LENGTH: 546
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-616-187-44

Query Match 99.7%; Score 2693; DB 15; Length 546;
Best Local Similarity 99.8%; Pred. No. 1.2e-155;
Matches 529; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 1 KSSPGQPEAGPEGAQERPSQAAPVAEAGPGSSQAAPRKEPGAQARTAQSGALRDVSEELS 60
Db 17 KSSPGQPEAGPEGAQERPSQAAPVAEAGPGSSQAAPRKEPGAQARTAQSGALRDVSEELS 76
QY 61 RQLEDILSTYCVDNQGGPDEGAQEPAPEDAEKSRITYVARGPEPTPVVYGEKPS 120
Db 77 RQLEDILSTYCVDNQGGPDEGAQEPAPEDAEKSRITYVARGPEPTPVVYGEKPS 136
QY 121 KGPNTTEIRQSDVGDHRRPQEKKAQKGLGKEITLLMQTLNTLSTPEEKLAALCKKY 180
Db 137 KGPNTTEIRQSDVGDHRRPQEKKAQKGLGKEITLLMQTLNTLSTPEEKLAALCKKY 196
QY 181 ABLEEHNSQOMKLLQKQSQOLVQKDHRLRGEHSKAVLARSKLSLCRELQRNRSIK 240
Db 197 ABLEEHNSQOMKLLQKQSQOLVQKDHRLRGEHSKAVLARSKLSLCRELQRNRSIK 256
QY 241 EGVQARAEERKEKVTSHFQVTLNDIQLQMEQHNRNSKLQRQENMELARLKKLIEQY 300
Db 257 EGVQARAEERKEKVTSHFQVTLNDIQLQMEQHNRNSKLQRQENMELARLKKLIEQY 316

QY 301 ELREEHIDKVPKHKDLOQOLVDKLOQAQEMLEAEERHOREKDFLLKEAVESQRMCELM 360
DB 317 ELREEHIDKVPKHKDLOQOLVDKLOQAQEMLEAEERHOREKDFLLKEAVESQRMCELM 376
QY 361 KQOETHLKQOLALYTERFEFEFQNTLSKSSSVFTTFKQEMEKMTKKIKKLEKETTMYRSRW 420
DB 377 KQOETHLKQOLALYTERFEFEFQNTLSKSSSVFTTFKQEMEKMTKKIKKLEKETTMYRSRW 436
QY 421 ESSNKALLEMAEKTVDKLEGLQVKIORLEKLCRALQTERNDLNKRVQDLSAGGQSSL 480
DB 437 ESSNKALLEMAEKTVDKLEGLQVKIORLEKLCRALQTERNDLNKRVQDLSAGGQSSL 496
QY 481 TDSGPERRPGGAQAPSSPRVTEAPCPGAPSTEASGQTGPQBPPTSARA 530
DB 497 TDSGPERRPGGAQAPSSPRVTEAPCPGAPSTEASGQTGPQBPPTSARA 546

RESULT 12
US-09-962-055-5
; Sequence 5, Application US/09962055
; Patent No. US20020052033A1
; GENERAL INFORMATION:
; APPLICANT: Lees, Ann M.
; Lees, Robert S.
; Law, Simon W.
; Arjona, Anibal A.
; TITLE OF INVENTION: NOVEL LOW DENSITY LIPOPROTEIN
; BINDING PROTEINS AND THEIR USES IN DIAGNOSING AND
; TREATING ATHEROSCLEROSIS
; NUMBER OF SEQUENCES: 42
; CORRESPONDENCE ADDRESS:
; ADDRESSES: Fish & Richardson P.C.
; STREET: 225 Franklin Street
; CITY: Boston
; STATE: MA
; COUNTRY: USA
; ZIP: 02110-2804
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/962,055
; FILING DATE: 24-Sep-2001
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/979,608
; FILING DATE: 26-NOV-1997
; APPLICATION NUMBER: US 60/031,930
; FILING DATE: 27-NOV-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: Myers, Louis
; REGISTRATION NUMBER: 35,965
; REFERENCE/DOCKET NUMBER: 10797-002001 (formerly 3983/59818)
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 617/542-5070
; TELEFAX: 617/542-8906
; INFORMATION FOR SEQ ID NO: 5:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 557 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; FRAGMENT TYPE: internal
; SEQUENCE DESCRIPTION: SEQ ID NO: 5:
US-09-962-055-5

Query Match 91.2%; Score 2465.5; DB 9; Length 557;
Best Local Similarity 90.8%; Pred. No. 8.3e-142;
Matches 492; Conservative 12; Mismatches 25; Indels 13; Gaps 3;
QY 1 KSSPQPEAGAGAQGRPQRPAPAREAG-ASSQAPPEGAQAKTAQPGALCDVSEBELS 60

DB 17 KSSPQPEAGAGAQGRPQRPAPAREAG-ASSQAPPEGAQAKTAQPGALCDVSEBELS 75
QY 61 RQLEDILSTYCVNDNNOGGPGEAGQEPAPEDAEKSRITYVARNGEPBP-TPVVYGEKEP 119
DB 76 RQLEDILSTYCVNDNNOGGPGEAGQEPAPEDAEKSRITYVARNGEPBP-TPVVYGEKEP 135
QY 120 SKGDDNTEIROSDEVGRDHRPQEKKAKGLGKEITLLMOTLNTLSTPEEKLAALCKK 179
DB 136 STAEFGTTEIRTSDEVGRDHRPQEKKAKGLGKEITLLMOTLNTLSTPEEKLAALCKK 195
QY 180 YAEELLEHRNSQOMKLLQKQSOQLVQEKDHLRGEHSAVLARSKLKESLCRELQHRNRL 239
DB 196 YAEELLEHRNSQOMKLLQKQSOQLVQEKDHLRGEHSAVLARSKLKESLCRELQHRNRL 255
QY 240 KEEGVORAREEKEKEKVTSHFQVTLNDIOLOMEOHNEHNSKLROENMELAEKLLKLEQ 299
DB 256 KEEGVORAREEKEKEKVTSHFQVTLNDIOLOMEOHNEHNSKLROENMELAEKLLKLEQ 315
QY 300 YELREEHIDKVPKHKDLOQOLVDKLOQAQEMLEAEERHOREKDFLLKEAVESQRMCEL 359
DB 316 YELREEHIDKVPKHKDLOQOLVDKLOQAQEMLEAEERHOREKDFLLKEAVESQRMCEL 375
QY 360 MKQOETHLKQOLALYTERFEFEFQNTLSKSSSVFTTFKQEMEKMTKKIKKLEKETTMYRSR 419
DB 376 MKQOETHLKQOLALYTERFEFEFQNTLSKSSSVFTTFKQEMEKMTKKIKKLEKETTMYRSR 435
QY 420 WESSNKALLEMAEKTVDKLEGLQVKIORLEKLCRALQTERNDLNKRVQDLSAGGQGS 479
DB 436 WESSNKALLEMAEKTVDKLEGLQVKIORLEKLCRALQTERNDLNKRVQDLSAGGQGP 495
QY 480 LTDSPERRP-----EGPQAQAPSSPRVTEAPCPGAPSTEASGQTGPQBPPTSARA 528
DB 496 VSDSGPERRPPEATTSKEQVGGPQAVPNSPRATDACCAGAPSTEASGQTGPQBPPTTA 555
QY 529 RA 530
DB 556 TA 557

RESULT 13
US-09-976-740-5
; Sequence 5, Application US/09976740
; Publication No. US20020194633A1
; GENERAL INFORMATION:
; APPLICANT: Lees, Ann M.
; APPLICANT: Lees, Robert S.
; APPLICANT: Law, Simon W.
; APPLICANT: Arjona, Anibal A.
; TITLE OF INVENTION: NOVEL LOW DENSITY LIPOPROTEIN BINDING
; PROTEINS AND THEIR USE IN DIAGNOSING AND TREATING
; ATHEROSCLEROSIS
; FILE REFERENCE: 10797-004001
; CURRENT APPLICATION NUMBER: US/09/976,740
; CURRENT FILING DATE: 2001-10-12
; PRIOR APPLICATION NUMBER: 09/616,289
; PRIOR FILING DATE: 2000-07-14
; PRIOR APPLICATION NUMBER: US 08/979,608
; PRIOR FILING DATE: 1997-11-26
; PRIOR APPLICATION NUMBER: US 60/031,930
; PRIOR FILING DATE: 1996-11-27
; PRIOR APPLICATION NUMBER: US 60/048,547
; PRIOR FILING DATE: 1997-06-03
; NUMBER OF SEQ ID NOS: 53
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO: 5
; LENGTH: 557
; TYPE: PRT
; ORGANISM: Oryctolagus cuniculus
US-09-976-740-5

Query Match 91.2%; Score 2465.5; DB 9; Length 557;
Best Local Similarity 90.8%; Pred. No. 8.3e-142;
Matches 492; Conservative 12; Mismatches 25; Indels 13; Gaps 3;

QY 1 KSSPGQPEAGPAGQERPSQAAPAVAEAGPGSSQAAPRKPEGAQAARTAQSGALDVSSELS 60
Db 17 KSSPGQPEAGAGQERPSQAAPAVAEAGPGSSQAAPRKPEGAQAARTAQSGALDVSSELS 75
QY 61 ROLEDILSTYCVNNQGGPGEAGQEPAPAEAGPDAEKSRVYVARGPEP-TPVVYGEKEP 119
Db 76 ROLEDILSTYCVNNQGGPGEAGQEPAPAEAGPDAEKSRVYVARGPEP-TPVVYGEKEP 135
QY 120 SKGDPNTEIROSDEVDGRDHRPQKKKAKGLGKEITLLMOTLNTLSTPEKLAALCKK 179
Db 136 SKAEPGTETIROSDEVDGRDHRPQKKKAKGLGKEITLLMOTLNTLSTPEKLAALCKK 195
QY 180 YAELEHNRNSOKMKLLQKQSQQLVQEKDHLRGHSHKAVLARSKLESICRELQHRNRL 239
Db 196 YAELEHNRNSOKMKLLQKQSQQLVQEKDHLRGHSHKAVLARSKLESICRELQHRNRL 255
QY 240 KEEGVQARAEERBEKEKVTSHFQVTLNDIQLQMEQHNRNSKLRQENMELABRLKGLIEQ 299
Db 256 KEEGVQARAEERBEKEKVTSHFQVTLNDIQLQMEQHNRNSKLRQENMELABRLKGLIEQ 315
QY 300 YELREHIDKVPKHDLQQLVDAKLQQAQEMLKEAERHOREKDFLLKEAVESQRMCEL 359
Db 316 YELREHIDKVPKHDLQQLVDAKLQQAQEMLKEAERHOREKDFLLKEAVESQRMCEL 375
QY 360 MKQETHLKQALALYTEKFEFQNTLSKSEVFTTFQEMEKMTKKIKLEKETMYRSR 419
Db 376 MKQETHLKQALALYTEKFEFQNTLSKSEVFTTFQEMEKMTKKIKLEKETMYRSR 435
QY 420 WESSKALLEMAEKTVDKLEGLQVKIQRLEKLCRALQTERNDLNKRVQDLSAGQGS 479
Db 436 WESSKALLEMAEKTVDKLEGLQVKIQRLEKLCRALQTERNDLNKRVQDLSAGQGS 495
QY 480 LTDSGPERP-----EGGAQAPSSPRVTPAFCYPCGAPSTASGOTGPQPTSA 528
Db 496 VSDSGPERPPEATTSKSQVEGGAQVNSPRATDASCAGAPSTASGOTGPQPTTA 555
QY 529 RA 530
Db 556 TA 557

RESULT 14

US-10-671-242-5
; Sequence 5, Application US/10671242
; Publication No. US2004004049A1
; GENERAL INFORMATION:
; APPLICANT: Lees, Ann M.
; APPLICANT: Lees, Robert S.
; APPLICANT: Law, Simon W.
; APPLICANT: Arjona, Anibal A.
; TITLE OF INVENTION: NOVEL LOW DENSITY LIPOPROTEIN BINDING
; TITLE OF INVENTION: PROTEINS AND THEIR USE IN DIAGNOSING AND TREATING
; TITLE OF INVENTION: ATHEROSCLEROSIS
; FILE REFERENCE: 10797-004001
; CURRENT APPLICATION NUMBER: US/10/671,242
; CURRENT FILING DATE: 2003-09-24
; PRIOR APPLICATION NUMBER: US/09/616,289
; PRIOR FILING DATE: 2000-07-14
; PRIOR APPLICATION NUMBER: US 09/517,849
; PRIOR FILING DATE: 2000-03-02
; PRIOR APPLICATION NUMBER: US 08/979,608
; PRIOR FILING DATE: 1997-11-26
; PRIOR APPLICATION NUMBER: US 60/031,930
; PRIOR FILING DATE: 1996-11-27
; PRIOR APPLICATION NUMBER: US 60/048,547
; PRIOR FILING DATE: 1997-06-03
; NUMBER OF SEQ ID NOS: 53
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 5
; LENGTH: 557
; TYPE: PRT
; ORGANISM: Oryctolagus cuniculus

US-10-671-242-5

Query Match 91.2%; Score 2465.5; DB 12; Length 557;
Best Local Similarity 90.8%; Pred. No. 8.3e-142;
Matches 492; Conservative 12; Mismatches 25; Indels 13; Gaps 3;
QY 1 KSSPGQPEAGPAGQERPSQAAPAVAEAGPGSSQAAPRKPEGAQAARTAQSGALDVSSELS 60
Db 17 KSSPGQPEAGAGQERPSQAAPAVAEAGPGSSQAAPRKPEGAQAARTAQSGALDVSSELS 75
QY 61 ROLEDILSTYCVNNQGGPGEAGQEPAPAEAGPDAEKSRVYVARGPEP-TPVVYGEKEP 119
Db 76 ROLEDILSTYCVNNQGGPGEAGQEPAPAEAGPDAEKSRVYVARGPEP-TPVVYGEKEP 135
QY 120 SKGDPNTEIROSDEVDGRDHRPQKKKAKGLGKEITLLMOTLNTLSTPEKLAALCKK 179
Db 136 SKAEPGTETIROSDEVDGRDHRPQKKKAKGLGKEITLLMOTLNTLSTPEKLAALCKK 195
QY 180 YAELEHNRNSOKMKLLQKQSQQLVQEKDHLRGHSHKAVLARSKLESICRELQHRNRL 239
Db 196 YAELEHNRNSOKMKLLQKQSQQLVQEKDHLRGHSHKAVLARSKLESICRELQHRNRL 255
QY 240 KEEGVQARAEERBEKEKVTSHFQVTLNDIQLQMEQHNRNSKLRQENMELABRLKGLIEQ 299
Db 256 KEEGVQARAEERBEKEKVTSHFQVTLNDIQLQMEQHNRNSKLRQENMELABRLKGLIEQ 315
QY 300 YELREHIDKVPKHDLQQLVDAKLQQAQEMLKEAERHOREKDFLLKEAVESQRMCEL 359
Db 316 YELREHIDKVPKHDLQQLVDAKLQQAQEMLKEAERHOREKDFLLKEAVESQRMCEL 375
QY 360 MKQETHLKQALALYTEKFEFQNTLSKSEVFTTFQEMEKMTKKIKLEKETMYRSR 419
Db 376 MKQETHLKQALALYTEKFEFQNTLSKSEVFTTFQEMEKMTKKIKLEKETMYRSR 435
QY 420 WESSKALLEMAEKTVDKLEGLQVKIQRLEKLCRALQTERNDLNKRVQDLSAGQGS 479
Db 436 WESSKALLEMAEKTVDKLEGLQVKIQRLEKLCRALQTERNDLNKRVQDLSAGQGS 495
QY 480 LTDSGPERP-----EGGAQAPSSPRVTPAFCYPCGAPSTASGOTGPQPTSA 528
Db 496 VSDSGPERPPEATTSKSQVEGGAQVNSPRATDASCAGAPSTASGOTGPQPTTA 555
QY 529 RA 530
Db 556 TA 557

RESULT 15

US-10-023-529-5
; Sequence 5, Application US/10023529
; Publication No. US20020129388A1
; GENERAL INFORMATION:
; APPLICANT: Lees, Ann M.
; APPLICANT: Lees, Robert S.
; APPLICANT: Law, Simon W.
; APPLICANT: Arjona, Anibal A.
; TITLE OF INVENTION: NOVEL LOW DENSITY LIPOPROTEIN BINDING
; TITLE OF INVENTION: PROTEINS AND THEIR USE IN DIAGNOSING AND TREATING
; TITLE OF INVENTION: ATHEROSCLEROSIS
; FILE REFERENCE: 10797-004001
; CURRENT APPLICATION NUMBER: US/10/023,529
; CURRENT FILING DATE: 2001-12-17
; PRIOR APPLICATION NUMBER: 09/616,289
; PRIOR FILING DATE: 2000-07-14
; PRIOR APPLICATION NUMBER: US 09/517,849
; PRIOR FILING DATE: 2000-03-02
; PRIOR APPLICATION NUMBER: US 08/979,608
; PRIOR FILING DATE: 1997-11-26
; PRIOR APPLICATION NUMBER: US 60/031,930
; PRIOR FILING DATE: 1996-11-27
; PRIOR APPLICATION NUMBER: US 60/048,547
; PRIOR FILING DATE: 1997-06-03
; NUMBER OF SEQ ID NOS: 53


```

; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 5
; LENGTH: 557
; TYPE: PRT
; ORGANISM: Oryctolagus cuniculus
US-10-023-529-5

Query Match          91.2%; Score 2465.5; DB 13; Length 557;
Best Local Similarity 90.8%; Pred. No. 8.3e-142;
Matches 492; Conservative 12; Mismatches 25; Indels 13; Gaps 3;

1 KSSPGQPEAGPEGAQERPSQAAPAEVAGPGSSQAAPKPEGAQAARTAQSGALRDVSEELS 60
17 KSSPGQPEAGPEGAQERPSQAAPAEVAGPGSSQAAPKPEGAQAARTAQSGALRDVSEELS 75
61 RQLEDILSTVCVNNQGGPDEGAGQEPAPEDAKSRRTYVARNGEPPEP-TPVYGEKEP 119
76 RQLEDILSTVCVNNQGGPDEGAGQEPAPEDAKSRRTYVARNGEPPEP-TPVYGEKEP 135
120 SKGDPNTEIRQSDEVDGRDHRPQEKKAKGLGKEITLLMOTLNTLSTPEEKLAALCKK 179
136 SKAEPGTEIRTSDEVDGRDHRPQEKKAKGLGKEITLLMOTLNTLSTPEEKLAALCKK 195
180 YAEILLEHRNSQKQMLLQKQSQLVQEKDHLRGEHSKAVLARSKLSLCRELQRHNSL 239
196 YAEILLEHRNSQKQMLLQKQSQLVQEKDHLRGEHSKAVLARSKLSLCRELQRHNSL 255
240 KEEGVORAREEERKEKVTSHFQVTLNDIQLQMEQHNRNSKLRQENMELAEKLLIEQ 299
256 KEEGVORAREEERKEKVTSHFQVTLNDIQLQMEQHNRNSKLRQENMELAEKLLIEQ 315
300 YELREEHIDKVFHKDLQQLQVDAKLQQAQEMLKEAERHOREKDFLLKEAVESQRMCEL 359
316 YELREEHIDKVFHKDLQQLQVDAKLQQAQEMLKEAERHOREKDFLLKEAVESQRMCEL 375
360 MKQETHLKOQALYTEKEFEFQNTLSKSSSVFTTFOEMEMKTKKIKLEKETTYRSR 419
420 WESSNKALLEMAEKIVADKELEGQVKIQRLEKLCRALQTERNDLNKRVQDLSAGQGS 479
436 WESSNKALLEMAEKIVADKELEGQVKIQRLEKLCRALQTERNDLNKRVQDLSAGQGS 495
480 LTDSGPERRP-----EGPGAQAPSSPRVTEAPCYGAPSTEASGQTGPQETSA 528
496 VSDSGPERRPPEPATTSKEQGVGPGAQVPSNPRATDASCCAGAPSTEASGQTGPQETSA 555

529 RA 530
556 TA 557
```

```

RESULT 17
US-10-616-187-5
; Sequence 5, Application US/10616187
; Publication No. US20040013668A1
; GENERAL INFORMATION:
; APPLICANT: Lees, Robert S.
; APPLICANT: Law, Simon W.
; APPLICANT: Arjona, Anibal A.
; TITLE OF INVENTION: NOVEL LOW DENSITY LIPOPROTEIN BINDING
; TITLE OF INVENTION: PROTEINS AND THEIR USE IN DIAGNOSING AND TREATING
; TITLE OF INVENTION: ATHEROSCLEROSIS
; FILE REFERENCE: 10797-004001
; CURRENT APPLICATION NUMBER: US/10/616,187
; CURRENT FILING DATE: 2003-07-09
; PRIOR APPLICATION NUMBER: US/09/616,289
```

```

; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 5
; LENGTH: 557
; TYPE: PRT
; ORGANISM: Oryctolagus cuniculus
US-10-023-529-5

Query Match          91.2%; Score 2465.5; DB 13; Length 557;
Best Local Similarity 90.8%; Pred. No. 8.3e-142;
Matches 492; Conservative 12; Mismatches 25; Indels 13; Gaps 3;

1 KSSPGQPEAGPEGAQERPSQAAPAEVAGPGSSQAAPKPEGAQAARTAQSGALRDVSEELS 60
17 KSSPGQPEAGPEGAQERPSQAAPAEVAGPGSSQAAPKPEGAQAARTAQSGALRDVSEELS 75
61 RQLEDILSTVCVNNQGGPDEGAGQEPAPEDAKSRRTYVARNGEPPEP-TPVYGEKEP 119
76 RQLEDILSTVCVNNQGGPDEGAGQEPAPEDAKSRRTYVARNGEPPEP-TPVYGEKEP 135
120 SKGDPNTEIRQSDEVDGRDHRPQEKKAKGLGKEITLLMOTLNTLSTPEEKLAALCKK 179
136 SKAEPGTEIRTSDEVDGRDHRPQEKKAKGLGKEITLLMOTLNTLSTPEEKLAALCKK 195
180 YAEILLEHRNSQKQMLLQKQSQLVQEKDHLRGEHSKAVLARSKLSLCRELQRHNSL 239
196 YAEILLEHRNSQKQMLLQKQSQLVQEKDHLRGEHSKAVLARSKLSLCRELQRHNSL 255
240 KEEGVORAREEERKEKVTSHFQVTLNDIQLQMEQHNRNSKLRQENMELAEKLLIEQ 299
256 KEEGVORAREEERKEKVTSHFQVTLNDIQLQMEQHNRNSKLRQENMELAEKLLIEQ 315
300 YELREEHIDKVFHKDLQQLQVDAKLQQAQEMLKEAERHOREKDFLLKEAVESQRMCEL 359
316 YELREEHIDKVFHKDLQQLQVDAKLQQAQEMLKEAERHOREKDFLLKEAVESQRMCEL 375
360 MKQETHLKOQALYTEKEFEFQNTLSKSSSVFTTFOEMEMKTKKIKLEKETTYRSR 419
420 WESSNKALLEMAEKIVADKELEGQVKIQRLEKLCRALQTERNDLNKRVQDLSAGQGS 479
436 WESSNKALLEMAEKIVADKELEGQVKIQRLEKLCRALQTERNDLNKRVQDLSAGQGS 495
480 LTDSGPERRP-----EGPGAQAPSSPRVTEAPCYGAPSTEASGQTGPQETSA 528
496 VSDSGPERRPPEPATTSKEQGVGPGAQVPSNPRATDASCCAGAPSTEASGQTGPQETSA 555

529 RA 530
556 TA 557
```

```

RESULT 16
US-10-023-523-5
; Sequence 5, Application US/10023523
; Publication No. US20020152485A1
; GENERAL INFORMATION:
; APPLICANT: Lees, Robert S.
; APPLICANT: Law, Simon W.
; APPLICANT: Arjona, Anibal A.
; TITLE OF INVENTION: NOVEL LOW DENSITY LIPOPROTEIN BINDING
; TITLE OF INVENTION: PROTEINS AND THEIR USE IN DIAGNOSING AND TREATING
; TITLE OF INVENTION: ATHEROSCLEROSIS
; FILE REFERENCE: 10797-004001
; CURRENT APPLICATION NUMBER: US/10/023,523
; CURRENT FILING DATE: 2001-12-17
; PRIOR APPLICATION NUMBER: US/09/616,289
; PRIOR FILING DATE: 2000-07-14
; PRIOR APPLICATION NUMBER: US 09/617,849
; PRIOR FILING DATE: 2000-03-02
; PRIOR APPLICATION NUMBER: US 08/979,608
; PRIOR FILING DATE: 1997-11-26
```

```

; PRIOR FILING DATE: 2000-07-14
; PRIOR APPLICATION NUMBER: US 09/517,849
; PRIOR FILING DATE: 2000-03-02
; PRIOR APPLICATION NUMBER: US 08/979,608
; PRIOR FILING DATE: 1997-11-26
; PRIOR APPLICATION NUMBER: US 60/031,930
; PRIOR FILING DATE: 1996-11-27
; PRIOR APPLICATION NUMBER: US 60/048,547
; PRIOR FILING DATE: 1997-06-03
; NUMBER OF SEQ ID NOS: 53
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 5
; LENGTH: 557
; TYPE: PRT
; ORGANISM: Oryctolagus cuniculus
; US-10-616-187-5

Query Match      91.2%; Score 2465.5; DB 15; Length 557;
Best Local Similarity 90.8%; Pred. No. 8.3e-142;
Matches 492; Conservative 12; Mismatches 25; Indels 13; Gaps 3;

Qy 1 KSSPGQPEAGPEGAQERPSQAAPAVEAGPGSSQAQPRKPEGAQARTAQSGALRDVSEELS 60
Db 17 KSSPGQPEAGPEGAQERPSQAAPAVEAGPGSSQAQPRKPEGAQARTAQSGALRDVSEELS 75
Qy 61 RQLEDILSTVCYDNNQGGPGEDGAQGEPAEPDAEKSRITYVARNGEPEP-TPVYVGEKEP 119
Db 76 RQLEDILSTVCYDNNQGGPGEDGVQGEPEPEPEPEPEPEPEPEPEPEPEPEPEPEPEPEPE 135
Qy 120 SKGDPNTEETROSDEVGDRDRHRRPQEKKAGLGKEITLLMOTINTLSTPEEKLAALCKK 179
Db 136 SKAEPGTEETROSDEVGDRDRHRRPQEKKAGLGKEITLLMOTINTLSTPEEKLAALCKK 195
Qy 180 YAEILLEHRNSOKMKLLQKKQSQLVQEKDHLRGEHSKAVLARSKLSLRELOHNRSL 239
Db 196 YAEILLEHRNSOKMKLLQKKQSQLVQEKDHLRGEHSKAVLARSKLSLRELOHNRSL 255
Qy 240 KEEGVQARABEEERKEVTSHTFQVTLNDIQLQMEQHNRNSKLRQENWELAEKLIIEQ 299
Db 256 KEEGVQARABEEERKEVTSHTFQVTLNDIQLQMEQHNRNSKLRQENWELAEKLIIEQ 315
Qy 300 YELREHIDKVFHKDLQQLVDKLAQQAQEMLKEAERHOREKDFLLKEAVESQRMCEL 359
Db 316 YELREHIDKVFHKDLQQLVDKLAQQAQEMLKEAERHOREKDFLLKEAVESQRMCEL 375
Qy 360 MKQETHLKKQALALYTEKFEFQNTLSKSSSEVFTTFKQEKEMTKKIKKLEKETTMYRSR 419
Db 376 MKQETHLKKQALALYTEKFEFQNTLSKSSSEVFTTFKQEKEMTKKIKKLEKETTMYRSR 435
Qy 420 WESSKALLEWAEBEKTVDKLEGLQVKIORLEKICRALQTERDNLKRVQDLGAGQGS 479
Db 436 WESSKALLEWAEBEKTVDKLEGLQVKIORLEKICRALQTERDNLKRVQDLGAGQGP 495
Qy 480 LTDSPERRP-----EGGAQAPSSPRVTEAPCYFGAPSTEASGQTGPQPTSA 528
Db 496 VSDSGPERRPATTSKQGVGPGCAQVNSPRATDASCCAGAPSTEASGQTGPQPTTA 555
Qy 529 RA 530
Db 556 TA 557

RESULT 18
US-10-276-774-2134
; Sequence 2134, Application US/10276774
; Publication No. US20040053245A1
; GENERAL INFORMATION:
; APPLICANT: Hysed, Inc.
; APPLICANT: Tang, Y. Tom et al
; TITLE OF INVENTION: No. US20040053245A1el Nucleic Acids and Polypeptides
; FILE REFERENCE: 21272-030
; CURRENT APPLICATION NUMBER: US/10/276,774
; CURRENT FILING DATE: 2002-11-18

```

```

; PRIOR APPLICATION NUMBER: 09/560,875
; PRIOR FILING DATE: 2000-04-27
; PRIOR APPLICATION NUMBER: 09/496,914
; PRIOR FILING DATE: 2000-02-03
; NUMBER OF SEQ ID NOS: 2700
; SOFTWARE: Custom
; SEQ ID NO 2134
; LENGTH: 510
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)...(510)
; OTHER INFORMATION: Xaa = any amino acid or nothing
; US-10-276-774-2134

Query Match      86.0%; Score 2324; DB 12; Length 510;
Best Local Similarity 99.8%; Pred. No. 3e-133;
Matches 459; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 KSSPGQPEAGPEGAQERPSQAAPAVEAGPGSSQAQPRKPEGAQARTAQSGALRDVSEELS 60
Db 28 KSSPGQPEAGPEGAQERPSQAAPAVEAGPGSSQAQPRKPEGAQARTAQSGALRDVSEELS 87
Qy 61 RQLEDILSTVCYDNNQGGPGEDGAQGEPAEPDAEKSRITYVARNGEPEP-TPVYVGEKEP 120
Db 88 RQLEDILSTVCYDNNQGGPGEDGAQGEPAEPDAEKSRITYVARNGEPEP-TPVYVGEKEP 147
Qy 121 KGDPTNTEETROSDEVGDRDRHRRPQEKKAGLGKEITLLMOTINTLSTPEEKLAALCKK 180
Db 148 KGDPTNTEETROSDEVGDRDRHRRPQEKKAGLGKEITLLMOTINTLSTPEEKLAALCKK 207
Qy 181 AELLEHRNSOKMKLLQKKQSQLVQEKDHLRGEHSKAVLARSKLSLRELOHNRSL 240
Db 208 AELLEHRNSOKMKLLQKKQSQLVQEKDHLRGEHSKAVLARSKLSLRELOHNRSL 267
Qy 241 EEGVQARABEEERKEVTSHTFQVTLNDIQLQMEQHNRNSKLRQENWELAEKLIIEQ 300
Db 268 EEGVQARABEEERKEVTSHTFQVTLNDIQLQMEQHNRNSKLRQENWELAEKLIIEQ 327
Qy 301 ELREHIDKVFHKDLQQLVDKLAQQAQEMLKEAERHOREKDFLLKEAVESQRMCEL 360
Db 328 ELREHIDKVFHKDLQQLVDKLAQQAQEMLKEAERHOREKDFLLKEAVESQRMCEL 387
Qy 361 KQOETHLKKQALALYTEKFEFQNTLSKSSSEVFTTFKQEKEMTKKIKKLEKETTMYRSR 420
Db 388 KQOETHLKKQALALYTEKFEFQNTLSKSSSEVFTTFKQEKEMTKKIKKLEKETTMYRSR 447
Qy 421 ESNKALLEWAEBEKTVDKLEGLQVKIORLEKICRALQ 460
Db 448 ESNKALLEWAEBEKTVDKLEGLQVKIORLEKICRALQ 487

RESULT 19
US-10-425-114-55188
; Sequence 55188, Application US/10425114
; Publication No. US2004003488A1
; GENERAL INFORMATION:
; APPLICANT: Liu, Jingdong
; APPLICANT: Zhou, Yihua
; APPLICANT: Kovalic, David K.
; APPLICANT: Screen, Steven E
; APPLICANT: Tabaska, Jack E
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated with
; FILE REFERENCE: 38-21(53313)B
; CURRENT APPLICATION NUMBER: US/10/425,114
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 73128
; SEQ ID NO 55188
; LENGTH: 437
; TYPE: PRT

```

```

Query Match      16.5%; Score 444.5; DB 12; Length 314;
Best Local Similarity 31.4%; Pred. No. 2.8e-19;
Matches 119; Conservative 71; Mismatches 11; Indels 77; Gaps 11;

QY      127 EEIQRSEVGDGRDRRRPQKKKAGLCKEITLLMQTTLATLSTPEKRLAALCKKYAELEE 186
          * * * * *
4 KENASRQVQ-----APSDD-----KGSSE-----PKRKVI-----K 30

```

187 HRNSQOMKLOKQKQSOVLQVHKGHRSKAVLARSKLSLCLRLQHRNSLKEGVOR 246
31 HSKLEKDKELF-----QLAQYHVKVAERDQALAVKORLESLSLCLRLQHRNSLKEGVOR 85
247 AREBEKREKVTSHFQVTLNDIQLOM-EQNNRNSKLQNNMELARLKLLEQYELREB 305
86 STEGONMRTSLSKFHAIKGVSQKLEQVERISQL-BENNTLRSLKLADQYNTIQ 144
306 HIDKVFHKLQLOQVDAKLOQAQEMLEKEAERHOREKOPFLKEAVESQRMCELMKQOET 365
145 KYAHQLEKKEKLELELADLRLOQHE--KAAQHTQWQ-----LYAEQVSQLM-TTEK 193
366 HLKQOLALYTERPEEFONTLSKSEVFTTFQEMERKMTKKIKLEKETMYRSRWESSNK 425
194 NLRLQSLADGERFQHDALSKSENFETTKQEMERKMSVINKLKENFLKGCNSDI 253
426 ALLEMAEKTVRKLEGLQVQKLEKLCRALQTERNDLNKRVQDLSAGGQSLDTS 485
254 AIVKLEERETKQKLEKQKLESLCTLOAERK----- 291
486 ERPEGFGAQPSPRYTE 504
292 ----QGFSASIPDAPSSQE 306

RESULT 22
US-10-369-493-5130
; Sequence 5130, Application US/10369493
; Publication No. US20030233675A1
; GENERAL INFORMATION:
; APPLICANT: Cao, Yongwei
; APPLICANT: Hinkle, Gregory J.
; APPLICANT: Slater, Steven C.
; APPLICANT: Goldman, Barry S.
; APPLICANT: Chen, Xianfeng
; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF
; FILE REFERENCE: 38-10(52052)B
; CURRENT APPLICATION NUMBER: US/10/369,493
; PRIOR FILING DATE: 2003-02-28
; PRIOR FILING DATE: 2002-02-21
; NUMBER OF SEQ ID NOS: 47374
; SEQ ID NO 5130
; LENGTH: 346
; TYPE: PRT
; ORGANISM: Caenorhabditis elegans
US-10-369-493-5130

Query Match 13.9%; Score 375; DB 15; Length 346;
Best Local Similarity 30.8%; Pred. No. 5.4e-15;
Matches 105; Conservative 69; Mismatches 123; Indels 44; Gaps 9;

150 KGLGKEITLQW-----TLNTL-STPE-EKLAALCHKVAELLEEHRNSOKMKLQKQ 201
3 KNPGLKLSIFNEGDEAALLKLEGVDAEKVKQKLIKLAEE--SEKQNAELKIKVLDYK 60
202 SQVQKHLRGHSHKAVL-----ARSKLSLCLRLQHRNSLKEGVORAREBEKRE 256
61 VVKVDTUTKLERNOILLRTEBSKSLRELRCGLQXANHQTRERACAKMKLEVERGL 120
257 VTSHFQVTLNDIQLOM-EQNNRNSKLQNNMELARLKLLEQYELREB-----LREE 305
121 AVEQLKVTLDIKETNAEGRSKSLAEADNKLSEKFEIGHQYERKMKVIDQIQKCK 180
306 HIDKVFHKLQLOQVDAKLOQAQEMLEKEAERHOREKOPFLKEAVESQRMCELMKQOET 365
181 YWEYGTQKLETKLTKLAKLESQVKKK-----GMEKDELAKIMLEETARVGALKTEK 236
366 HLKQOLALYTERPEEFONTLSKSEVFTTFQEMERKMTKKIKLEKETMYRSRWESSNK 425
237 ALREQVEYSKSELTSLSKSNKNEAFDKFKDELSRVNKKCMQVKEGLSKKKSDBANK 296

426 ALL-----EMAEKTVRDKLEGLQVQKLEKLCRALQ 459
297 KVLVTWNTNOEVAEKIATSDK-----KIQMLENLCLRALR 330

RESULT 23
US-10-425-114-45419
; Sequence 45419, Application US/10425114
; Publication No. US20040034888A1
; GENERAL INFORMATION:
; APPLICANT: Liu, Jingdong
; APPLICANT: Zhou, Yihua
; APPLICANT: Kovalic, David K.
; APPLICANT: Screen, Steven E.
; APPLICANT: Tabaska, Jack E.
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated with
; FILE REFERENCE: 38-21(53313)B
; CURRENT APPLICATION NUMBER: US/10/425,114
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 73128
; SEQ ID NO 45419
; LENGTH: 206
; TYPE: PRT
; ORGANISM: Zea mays
; FEATURE:
; OTHER INFORMATION: Clone ID: LIB3088-030-A1_FLI.pep
US-10-425-114-45419

Query Match 11.3%; Score 304.5; DB 12; Length 206;
Best Local Similarity 34.5%; Pred. No. 5.5e-11;
Matches 78; Conservative 48; Mismatches 77; Indels 23; Gaps 6;

273 EQNNRNSKLQNNMELARLKLLEQYELREBHDKVKHKLQLOQVDAKLOQAQEML 332
4 EQVECTIAQLENNM-LRSKLDIADQYNTIQQYAHQLEKQMELELADLRLOQHE-- 60
333 KEAERHOREKOPFLKEAVESQRMCELMKQOETHLQKQALALYTERPEEFONTLSKSEV 392
61 KAAHEHTQWQ-----LYAEQVSQLM-TTEKRLQLASGERFQHDALSKSEV 111
393 TTFQEMERKMTKKIKLEKETMYRSRWESSNKALLEMAEKTVRDKLEGLQVQK 452
112 ETVKQEMERKMLVINKLKDNEFLKGCNSDIALVKLIERELTKQTEKLNQKETTLE 171
453 KLCRALQTERNDLNKRVQDLSAGGQSLDTSQRPERRPBGPGAQPS 498
172 SLCSLQAEKQ-----GRSASIPDA-PSSQEDMPATSQES 206

RESULT 24
US-10-369-493-1016
; Sequence 1016, Application US/10369493
; Publication No. US20030233675A1
; GENERAL INFORMATION:
; APPLICANT: Cao, Yongwei
; APPLICANT: Hinkle, Gregory J.
; APPLICANT: Slater, Steven C.
; APPLICANT: Goldman, Barry S.
; APPLICANT: Chen, Xianfeng
; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF
; FILE REFERENCE: 38-10(52052)B
; CURRENT APPLICATION NUMBER: US/10/369,493
; CURRENT FILING DATE: 2003-02-28
; PRIOR FILING DATE: 2002-02-21
; NUMBER OF SEQ ID NOS: 47374
; SEQ ID NO 1016
; LENGTH: 886
; TYPE: PRT

```

; ORGANISM: Archaeoglobus fulgidus
US-10-369-493-1016

Query Match      9.6%; Score 259; DB 15; Length 886;
Best Local Similarity 24.8%; Pred. No. 1.9e-07;
Matches 113; Conservative 86; Mismatches 111; Indels 146; Gaps 22;

QY 116 EKPSKDPNTEETIRQSDVGRDHRPPOEKKAKGLKETTLLMOTLNTLSTPEEKLA 175
DB 171 EKEELK-----EFLSQEQL-----KQKEKKK-----EERISBEIKSISREKLS- 214
QY 176 LCKXYAELLEHRNSQOMKLLQKQSQOLVQKDHGRGSHKAVIARSKLESICREIQRH 235
DB 215 -----BEVRNLSRLKESEHKSRL-----ESLAKQESSVLQVGRGLEEKLEBLE-- 259
QY 236 NRSLESGVQPARSEEEKKEV-----TSHFQVTLND----- 267
DB 260 -KQKEV-VERIELEKKAKEVKEKPKAERYSTLEKLLSEINQALRDVKEKRGDLTRE 317
QY 268 --IQLOMEQHNRNSKLQENMELAEKLLKLIQYE----- 303
DB 318 AGIQALKAEDNSKL-BEITRKIEBLERLEFERFKSHRLLETLPKMDRMQGIKAKLE 376
QY 304 EEHT--DKVPKHQLOQOLVDKALQQAQEMLKEABERHOR-----EKDFL-----LKEAVE 352
DB 377 EKNUTPKVEQWYDL-----LSKAKEBEKEITEKLLIAKSSLKTRGAQLKCAVE 428
QY 353 ----SORMCELMQO--ETHLKQOLALYTRKPFQNTLSKSSSEVFTTFKQEMKMTKKI 406
DB 429 ELKSAERTCPVCGRELDSEHRKNIAEYTRMKIAESLAKADEIEKKLERLEKVE--- 485
QY 407 KLEKETMTWRSRH-----ESNKALLEMAEKTVTRDKLEGL--QVKI-- 448
DB 486 KALEKQETVLKRYQWVDELKALENELSHDAEKLSAESEYRYKYKER-LDGLRQOQKILL 544
QY 449 -----QRLKLCRALQTERNDLNKRVOD 471
DB 545 SSASRIKELSSLSREIEALKNWSESGELRRKIRE 580

RESULT 25
US-10-072-012-838
; Sequence 838, Application US/10072012
; Publication No. US20040033493A1
; GENERAL INFORMATION:
; APPLICANT: Tchernev, Velizar
; APPLICANT: Spytek, Kimberly
; APPLICANT: Zethusen, Bryan
; APPLICANT: Patturajan, Meera
; APPLICANT: Shimkets, Richard
; APPLICANT: Li, Li
; APPLICANT: Gangolli, Beha
; APPLICANT: Padigaru, Muralidhara
; APPLICANT: Anderson, David W.
; APPLICANT: Rastelli, Luca
; APPLICANT: Miller, Charles E.
; APPLICANT: Gerlach, Valerie
; APPLICANT: Taupier Jr, Raymond J.
; APPLICANT: Gusev, Vladimir Y.
; APPLICANT: Colman, Steven D.
; APPLICANT: Wolenc, Adam R.
; APPLICANT: Pena, Carol E. A
; APPLICANT: Furtak, Katarzyna
; APPLICANT: Grosse, William M.
; APPLICANT: Alsobrook II, John P.
; APPLICANT: Lepley, Denise M.
; APPLICANT: Rieger, Daniel K.
; APPLICANT: Burgess, Catherine E.
; TITLE OF INVENTION: Proteins and Nucleic Acids Encoding Same
; FILE REFERENCES: 21402-258
; CURRENT APPLICATION NUMBER: US/10/072,012
; CURRENT FILING DATE: 2002-01-31
; PRIOR APPLICATION NUMBER: 60/265,102

; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: 60/265,514
; PRIOR FILING DATE: 2001-01-31
; PRIOR APPLICATION NUMBER: 60/265,517
; PRIOR FILING DATE: 2001-01-31
; PRIOR APPLICATION NUMBER: 60/265,412
; PRIOR FILING DATE: 2001-01-31
; PRIOR APPLICATION NUMBER: 60/265,395
; PRIOR FILING DATE: 2001-01-31
; PRIOR APPLICATION NUMBER: 60/266,406
; PRIOR FILING DATE: 2001-02-02
; PRIOR APPLICATION NUMBER: 60/266,767
; PRIOR FILING DATE: 2001-02-05
; PRIOR APPLICATION NUMBER: 60/267,057
; PRIOR FILING DATE: 2001-02-07
; PRIOR APPLICATION NUMBER: 60/266,975
; PRIOR FILING DATE: 2001-02-07
; PRIOR APPLICATION NUMBER: 60/267,459
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 1391
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 838
; LENGTH: 860
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: Myosin tail
US-10-072-012-838

Query Match      9.5%; Score 257.5; DB 12; Length 860;
Best Local Similarity 22.7%; Pred. No. 2.3e-07;
Matches 143; Conservative 106; Mismatches 189; Indels 193; Gaps 25;

QY 12 EGAQRPSQAAPAAVAAEGPSSQAPRKDEGAQARTAQSGALRDVGEELSRLQEDILSTYC 71
DB 77 EELSERLEAGGATAAQ-----IELNKKRAELAKLAKDLLEANIQHEALAT-L 125
QY 72 VDNQGGPEDGAQGPAPBPDAEKSRITYVANGPEPPTPVVYGEKPKSGDPNTEE-IR 130
DB 126 RKKHQDAINELSEQIEQLQKQAK-----AEKBSQLAQAEVDDLLA 166
QY 131 QSDVEGDRDHRPPOEKKAKGLKBIT-----LIMOTLNTLSTPEEKLA-- 175
DB 167 QLDSE---TKAKLNAEKKAKQLESQLESLQVLEKLDLQNLDTQKSELQSENLDLTFQ 223
QY 176 -----LCKYABL---LEEHRNS-----QKQMLLKQSQQLVQEKHLRGHSA 218
DB 224 LEEBAQVSNLSKLKSLQLESQLEAKRSLEESRERANLQALRQLEHDLDSLREQLEEE 283
QY 219 VLARSKLSLCREIQRHNRSLKE-----EGVQPARSEEEKRKEVT----- 258
DB 284 SEAKAELE---RQLSKANAEIQWRSKFSEGAURAELEELKLLNOKISLEBAEAA 340
QY 259 -----SHFQVTLNDIQLOMEQHNRNSKL--RQENME--LAERLKKLIE----- 298
DB 341 NAKCDSLEKTKSRLQSELEDLQILERANAAASELEKQKNPDKILAERKRVDELQABL 400
QY 299 ---OYELR-----EEHDKV---FKHKDLOQQLVDKALQQAQ-----EMLK 333
DB 401 DTAQREARNLSTELPRLKNELEFLKQVQVEALRRKNLQDEIHDLTDQLGEGGRVWHELE 460
QY 334 EAEERHQEKDFLLKEAVESQRMCEL-----MKQOETHLKQOLALYTEKPEEFQ 382
DB 461 KARRLEAKKDELQALAEARAALEESKVLRAQVELSQIRSEIERLLA---EKESEFE 517
QY 383 NT-----LSKSSSEVFTTFKQEMKMTKK-----IK 407
DB 518 NTRKNHORAISLQATLEATKGAESRLKKKLEGGDINELEIALDHANKANAEAKNNVK 577
QY 408 KLEKETMTWRSWESSNKA-----LLEWABEK-TVRDKLEGLQVKTORLEKLCRALQT 460
DB 578 KYQQQVKELQTVBEQPARDAEQALAVARRATALAELEELRSALSQAERARKQAE 637
```

QY	461	ERNDLNKRVQDLSAGGQGSILTSGPERRPREG	491
Db	638	ELAEASERVNLTQNSSLIAQ---KKLEG 665	
RESULT 26			
US-10-037-417-59			
; Sequence 59, Application US/10037417			
; Publication No. US20040052806A1			
; GENERAL INFORMATION:			
; APPLICANT: Kekuda, Ramesh			
; APPLICANT: Alsobrook II, John P			
; APPLICANT: Tchernev, Velizar T			
; APPLICANT: Liu, Xiaohong			
; APPLICANT: Spytek, Kimberly A			
; APPLICANT: Patturajan, Meera			
; APPLICANT: Grosse, William M			
; APPLICANT: Lepley, Denise M			
; APPLICANT: Burgess, Catherine E			
; APPLICANT: Vernet, Corine A.M.			
; APPLICANT: Li, Li			
; APPLICANT: Gorman, Linda			
; APPLICANT: Edinger, Shlomit R			
; APPLICANT: Sciore, Paul			
; APPLICANT: Ellerman, Karen			
; APPLICANT: Malyankar, Uriel M			
; APPLICANT: Rotherberg, Mark			
; APPLICANT: Stone, David J			
; APPLICANT: Boldog, Ferenc L			
; APPLICANT: Guo, Xiaojia			
; APPLICANT: Shenoy, Suresh G			
; APPLICANT: Anderson, David W			
; APPLICANT: Padigaru, Muralidhara			
; APPLICANT: Taupier Jr, Raymond J			
; APPLICANT: Miller, Charles E			
; APPLICANT: Eisen, Andrew J			
; TITLE OF INVENTION: Proteins and Nucleic Acids Encoding Same			
; FILE REFERENCE: 21402-235			
; CURRENT APPLICATION NUMBER: US/10/037,417			
; CURRENT FILING DATE: 2002-09-20			
; PRIOR FILING DATE: 2002-09-20			
; PRIOR FILING DATE: 2001-01-05			
; PRIOR FILING DATE: 2001-01-05			
; PRIOR FILING DATE: 2001-01-08			
; PRIOR FILING DATE: 2001-01-08			
; PRIOR FILING DATE: 2001-02-28			
; PRIOR FILING DATE: 2001-02-28			
; PRIOR FILING DATE: 2001-03-02			
; PRIOR FILING DATE: 2001-03-02			
; PRIOR FILING DATE: 2001-05-15			
; PRIOR FILING DATE: 2001-05-15			
; PRIOR FILING DATE: 2001-07-05			
; PRIOR FILING DATE: 2001-07-05			
; PRIOR FILING DATE: 2001-07-12			
; PRIOR FILING DATE: 2001-07-12			
; PRIOR FILING DATE: 2001-09-10			
; PRIOR FILING DATE: 2001-09-10			
; PRIOR FILING DATE: 2001-09-12			
; NUMBER OF SEQ ID NOS: 227			
; SOFTWARE: PatentIn Ver. 2.1			
; SEQ ID NO 59			
; LENGTH: 860			
; TYPE: PRT			
; ORGANISM: Artificial Sequence			
; FEATURE:			
; OTHER INFORMATION: Description of Artificial Sequence: Myosoin Tail			
; OTHER INFORMATION: Consensus Sequence			
US-10-037-417-59			
Query Match 9.5%; Score 257.5; DB 12; Length 860;			
Best local Similarity 22.7%; Pred. No. 2.3e-07;			
Matches 143; Conservative 106; Mismatches 189; Indels 193; Gaps 25;			

QY	12	EGAOERPSQAAPVAREGPGSSQAPRKPEGAQARTAQSGALRDVSELSRQLEDILSTVC	71
Db	77	BELSERLEBEAGGATAQ-----JELNKRREAEELAKRLKOLEEALNQEHEALAT-L	125
QY	72	VNNQGGCHDGAQCHPEAPEDAEKSRVVAANGPEPTPVVYGEKESKCDPNTTE-IR	130
Db	126	RKKHGDAINELSEIQEQKQKAK-----AEKESQLAEVDDLLA	166
QY	131	QSDVEYDGRHRRPQEKKAAGLGKEIT-----ILLQTLNTLSTPPEEKLA-	175
Db	167	OLDSI---TKAKLNAEKAKQLESQLESLQVVKLDELQRLNDLTSQKSLQSENSDLTRQ	223
QY	176	-----LCKYVABL---LEHRNS-----OKQKLLQKQSQVQVQKHLRGHSKA	218
Db	224	LEEAQVSNLSKLSQLESQLEEAQRSLEBSREANLQALQRLDLDLSRQLEEE	283
QY	219	VLARSKLSLCLRELQHRNSLKE-----EGVQBARBEERKKRKEVT-----	258
Db	284	SEAKAELE---ROLSKANAEIQOWSKFSESEGAJAELELKKLNQKISLEBAABAA	340
QY	259	-----SHFQVTLNDIQVQVQHNERNKIL--ROENME--LAERLKKLIE-----	298
Db	341	NAKCDSEKTKSRQSELEDLQIELERANAAASELEKQKQKQKFDKILAEWKRVDBLOAEL	400
QY	299	---QYELR-----SEHIDKV---FKHDLQOQLVDAKLQQAQ-----EMLK	333
Db	401	DTAQREARNLSTELFRLKVELSELKQVQVEALPREKNLQDEIHDLTDQLGEGRNWHELE	460
QY	334	EAEERHOREKDFLLKEAVESQRMCEL-----MKQQETHLKKQALVYTKFEPFQ	382
Db	461	KARRLEAEKDELQAALBEARAALEBSKVLRQAQVELSQIRSEIERLIA---EKESPE	517
QY	383	NT-----LSKSSSVFTTFKQEMKMTKK-----LK	407
Db	518	NTRKNHORAIBLSQATLEATYKGAASRLKKKLGSDINELEIALDHANKANAQKQVVK	577
QY	408	KLEKETTVYRSWESSNKA-----LLEVAERK-TVRDKLEGLQVVKIQRLKLCRALQT	460
Db	578	KYQQQVGEIQTQVEBEQARERDAREQLAVAEARRATALEAELEBSALSALQOABRARKQAST	637
QY	461	ERNDLNKRVQDLSAGGQGSILTSGPERRPREG	491
Db	638	ELAEASERVNLTQNSSLIAQ---KKLEG 665	

RESULT 27

US-10-080-334-166

; Sequence 166, Application US/10080334

; Publication No. US20040002584A1

; GENERAL INFORMATION:

; APPLICANT: Pena, Carol E. A.

; APPLICANT: Shimkets, Richard A

; APPLICANT: Li, Li

; APPLICANT: Shenoy, Suresh G

; APPLICANT: Kekuda, Ramesh

; APPLICANT: Spytek, Kimberly A

; APPLICANT: Vernet, Corine A. M.

; APPLICANT: Malyankar, Uriel M

; APPLICANT: Guo, Xiaojia

; APPLICANT: Gusev, Vladimir Y

; APPLICANT: Casman, Stacie J

; APPLICANT: Boldog, Ferenc L

; APPLICANT: Furtak, Katarzyna

; APPLICANT: Tchernev, Velizar T

; APPLICANT: Patturajan, Meera

; APPLICANT: Gangolli, Bsha A

; APPLICANT: Padigaru, Muralidhara

; APPLICANT: Liu, Xiaohong

; APPLICANT: Baumgartner, Jason C.

; APPLICANT: Gerlach, Valerie

; APPLICANT: Spaderna, Steven X

; APPLICANT: Zerhusen, Bryan D

QY 299 ---QVELR-----BEHIDKV-----FKHDLQOQLVDAKLOAQ-----EMLK 333
Db 401 DTAQREARNLSTELFLANELEELKQVFEALRRNNKQDIBHDLTDQGGGRNVHELE 460
QY 334 EAEERHOREKDFILKEAVESQWMBEL-----MKQOETHLKKOOLALYTEKPERFQ 382
Db 461 KARRLEBAKDELQALAEBALEAEKVLRAQVELSQIRSEIRRLA-----EKEEFBE 517
QY 383 NT-----LSKSEVFTFFQEMEMTK-----IK 407
Db 518 NTRKHORAIRESLOATLEAETKGAENSLRKKGLGEGDINELEIALDHANKANAERAKQK 577
QY 408 KLEKETTYRSHWESSNKA-----LLEWABEK-TVRDKELEGLOVKIQRLKLCALQ 460
Db 578 KYQQVKELQVVEEQEQAEDAREQOLAVERRATALEAELELSALEQAERARKQABT 637
QY 461 ERNDLKNVQDLSAGGQSLTDSGPERPBG 491
Db 638 ELAEASERVNLTACNSSLIAQ---KRLKLG 665

RESULT 28
US-10-146-473-41
; Sequence 41, Application US/10146473
; Publication NO. US2003010888A1
; GENERAL INFORMATION:
; APPLICANT: Scanlan, Matthew
; APPLICANT: Gout, Ivan
; APPLICANT: Stockert, Elisabeth
; APPLICANT: Gure, Ali
; APPLICANT: Chen, Yao-Tseng
; APPLICANT: Old, Lloyd
; TITLE OF INVENTION: Breast Cancer Antigens
; FILE REFERENCE: L00461/70130(JRV)
; CURRENT APPLICATION NUMBER: US/10/146,473
; CURRENT FILING DATE: 2002-05-15
; PRIOR APPLICATION NUMBER: US 60/291,150
; PRIOR FILING DATE: 2001-05-15
; NUMBER OF SEQ ID NOS: 82
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 41
; LENGTH: 2871
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-146-473-41

Query Match 9.5%; Score 256; DB 14; Length 2871;
Best Local Similarity 23.2%; Pred. No. 1.2e-06;
Matches 133; Conservative 107; Mismatches 195; Indels 138; Gaps 25;

QY 26 EAGPGSSQAPKXP-----EQAQARTAS-----GALRDVSELSRQLEDIL 67
Db 1432 QKATGSEVSQKQLEVLRLQVMTQRTSESVYKOSLDDAKTIQDKKEIER-LKQLI 1490
QY 68 STYCVNNQGGEDGAQGEPAEPDAEKSR-TYVARNGEPPFPVY-----YGEKPSKG 122
Db 1491 DKETNDR-----KLEDEANARLQVQYDLQKANSATETINKLKVOEQLTRL 1538
QY 123 DPWTEIRQSDVGRDRHRPQEKKAQGLKEITLMTLNTLSTPEBKALALCKYAE 182
Db 1539 RIDYERVSQRTYKQDITRFQNSLKELQKQKQ-KVEELNKLKRTASDSCRRKLEB 1596
QY 183 LLEHNRNSQKMKL-----LQKQSQ--LVQEKD-----HLR-GEHSAVLAR 222
Db 1597 ELGWRSLKEQAKITNLTLQQLSQASIVKRSDDLQQRDVLGDHLREKQRTQELRR 1656
QY 223 --SKLESICHELQHRNLSKE-----EGVQARBEERKEKVTSHQVTLNDIQLQWEQH 275
Db 1657 LSSEVALRQLIQBQSVKQAHLEHFKQAKEDKSR-----SUNESKIEIRL 1706
QY 276 NERNKLRQENMELAEKRLK-LTEQVELR-----EZHDK-----VFKHDLQ 318
Db 1707 QSLTENLTKEHMLLEELNRLLEYDDLRGRSADSKNATILELSQISNNRTLEL 1766

QY 12 EGAQERPSQAAPAVEAGFGSQAPRKPGEQAQRTAQSCALRDVSELSRQLEDILSTYC 71
Db 77 EELSRLBEAGGATAAQ-----IELNKGAEALAKRLKDLBEANLQHEALAT-L 125
QY 72 VDNQGGPGEDGAQGEPAEPDAEKSR-TYVARNGEPEPTVYVYGEKPSKGDPTHE-IR 130
Db 126 RKXHQDALNELSEQLQKQKAK-----AEKESQLOAEVDDLLA 166
QY 131 QSDVGRDRHRPQEKKAQGLKEIT-----LLMOTLNTLSTPEBKALAA----- 175
Db 167 QLDSI---TKAKLNAAEKKAQLESQSLQVLDLQRLNDLTSQKRSQSENSDLTRQ 223
QY 176 -----LCKYAEGL-----LLEHNRNS-----QKMKLQKQSQLOVQEKDLRGEHSA 218
Db 224 LEEAQAQVNLKLSQLESQLEAEKRLSESRERANLQQLRQLEHDLQSLAEQLEEB 283
QY 219 VLARSLKLSLCHLQHRNLSKE-----EGVQARBEERKEKVT----- 258
Db 284 SEAKAELE---ROLSKANAELIQWRSKFSEGALEAELEELKKNQKINSELBEAEAA 340
QY 259 -----SHFQVTLNDIQLQWEQHNRNLSK-L--ROENNE--LAERLKLILB----- 298
Db 341 NAKCDSLENTKSRLOSELEDIQLIELERANAAASELEKKQKQDKILAEWKRVDELQAE 400

Query Match 9.5%; Score 257.5; DB 15; Length 860;
Best Local Similarity 22.7%; Pred. No. 2.3e-07;
Matches 143; Conservative 106; Mismatches 189; Indels 193; Gaps 25;

US-10-080-334-166

TITLE OF INVENTION: Proteins, Polynucleotides Encoding Them and Methods of
FILE REFERENCE: 21402-275
CURRENT APPLICATION NUMBER: US/10/080,334
CURRENT FILING DATE: 2002-02-21
PRIOR APPLICATION NUMBER: 60/270,523
PRIOR FILING DATE: 2001-02-21
PRIOR APPLICATION NUMBER: 60/322,712
PRIOR FILING DATE: 2001-09-17
PRIOR APPLICATION NUMBER: 60/311,980
PRIOR FILING DATE: 2001-08-13
PRIOR APPLICATION NUMBER: 60/330,307
PRIOR FILING DATE: 2001-10-18
PRIOR APPLICATION NUMBER: 60/278,796
PRIOR FILING DATE: 2001-03-26
PRIOR APPLICATION NUMBER: 60/281,521
PRIOR FILING DATE: 2001-04-04
PRIOR APPLICATION NUMBER: 60/276,677
PRIOR FILING DATE: 2001-03-16
PRIOR APPLICATION NUMBER: 60/311,595
PRIOR FILING DATE: 2001-08-10
PRIOR APPLICATION NUMBER: 60/270,220
PRIOR FILING DATE: 2001-02-21
PRIOR APPLICATION NUMBER: 60/274,295
PRIOR FILING DATE: 2001-03-08
PRIOR APPLICATION NUMBER: 60/318,526
PRIOR FILING DATE: 2001-09-10
PRIOR APPLICATION NUMBER: 60/286,548
PRIOR FILING DATE: 2001-04-25
PRIOR APPLICATION NUMBER: 60/291,765
PRIOR FILING DATE: 2001-05-17
PRIOR APPLICATION NUMBER: 60/270,797
PRIOR FILING DATE: 2001-02-23
PRIOR APPLICATION NUMBER: 60/276,400
PRIOR FILING DATE: 2001-03-16
PRIOR APPLICATION NUMBER: 60/270,810
PRIOR FILING DATE: 2001-02-23
NUMBER OF SEQ ID NOS: 388
SOFTWARE: PatentIn Ver. 2.1
SEQ ID NO 166
LENGTH: 860
TYPE: PRT
ORGANISM: Homo sapiens
US-10-080-334-166

US-09-291-417-23
; Publication No. US020030050230A1
; Sequence 23, Application US/09291417A
; Publication No. US020030050230A1
; GENERAL INFORMATION:
; APPLICANT: FLOWMAN, GREGORY
; APPLICANT: MARTINEZ, RICARDO
; APPLICANT: WHYTE, DAVID
; TITLE OF INVENTION: STE20-RELATED PROTEIN KINASES
; FILE REFERENCE: 240/300
; CURRENT APPLICATION NUMBER: US/09/291.417A
; EARLIER FILING DATE: 1999-04-13
; EARLIER APPLICATION NUMBER: US 60/081,784
; EARLIER FILING DATE: 1998-04-14
; NUMBER OF SEQ ID NOS: 147
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 23
; LENGTH: 786
; TYPE: PRT
; ORGANISM: Mammalian (Human) SULU3
US-09-291-417-23

Query Match 9.2%; Score 248.5; DB 10; Length 786;
Best Local Similarity 21.6%; Pred. No. 7.2e-07;
Matches 139; Conservative 117; Mismatches 197; Indels 191; Gaps 29;

Qy 23 PAVAR-----GPGSSQA-PRKPEGAQAARTAGSALRDVSELSRLQLE 64
Db 109 PAVARQEBEEDHGVGRGTGVNSVGSNQSIPMSISASSOSSVNSLPDVSDDKS-BLD 167
Qy 65 DILSTYCVNNQG-----GPGSDG--AOGEPAPBDAKSTYTVARNGEPEPTPVVYGEKE 118
Db 168 MMEGDHTVMSNSSVIHLKPEENYREGDP-----RT---RASDPQSPQVSRHS 215
Qy 119 PSKGDPTTEIRQSDVEGDRHRRPOEKKAKGLGKEIT---LLMQTLNTLSTPEKLA 174
Db 216 HYRNRHFATIRTSLV---TRQMBEHDQSELRQMSGYKMRROHQKQMTLENK 271
Qy 175 ALCKKYA---ELLEHRN---SQQMKLLQKQSQLVQE-----KDLRGHSHK 217
Db 272 AEMDEHRLDKDLETQNNFAEMKLIKKHQAAMEKAKVMSNEKKFOHIAQOQKK 331
Qy 218 AVLARKSLKSLCRLQHRNRSKKEGVOR-----AREBEE---KKEVTSHPV--TLNDI 268
Db 332 EL--NSFLESQREYKLRKEQLKEELNENQSTPKKEQEWLSKQENIQHFOABEENLL 389
Qy 269 QLOME-----QHN-----ERNSKLRQENMELAEKLLRQYELRE 304
Db 330 PRQRYLELCERPKRMMLGRHNLQDLVREELNKRQTKDLEHMLLRQHSQMLEP 449
Qy 305 EHDVKFKHDLQOOLVDAKLQAQAEMLKEAEERHOREKDFLLKEAVESQRMCELMKQOE 364
Db 450 RHLNTIQK---MRCELI---RLQHOTELTNQLEYNKRERELRKHVMEVRRQPKSLKSKE 504
Qy 365 THLKQOL---ALYTEKEEFONTLSKSSSEVFTTFKQEMEKMTKKIKKLEKETTWYSRW 420
Db 505 LQIKKQFODTKIOTQRYKALRNHLE---TTPKSEHKAVLRKLK--EBQTRKLAILA 557
Qy 421 ESSNKALLEMAEBKTVR-----DKELE 442
Db 558 EYDHSINEMLSQALRLDQAEACQVLKMQLOQHELELLNAYQSKIKMQAEACHDREL 617
Qy 443 GLOVKI-----ORLEKLCRALQTERND-----LNKRVODLSAGGGS-----L 480
Db 618 ELEQVSLRRALLLEOKIEEMALQWTERIRSLERQARETEAFDSESRMLGFSNMVL 677
Qy 481 TDSGPERPEPGGAQAPSSPRVTEAPCYGAP--STEASGQTGP 522
Db 678 SNLSPE-----AFSHSYFGASGWSHNPTGGPGP 705

RESULT 32
US-09-291-417-31
; Sequence 31, Application US/09291417A

Publication No. US20030050230A1
; GENERAL INFORMATION:
; APPLICANT: FLOWMAN, GREGORY
; APPLICANT: MARTINEZ, RICARDO
; APPLICANT: WHYTE, DAVID
; TITLE OF INVENTION: STE20-RELATED PROTEIN KINASES
; FILE REFERENCE: 240/300
; CURRENT APPLICATION NUMBER: US/09/291.417A
; EARLIER FILING DATE: 1999-04-13
; EARLIER APPLICATION NUMBER: US 60/081,784
; EARLIER FILING DATE: 1998-04-14
; NUMBER OF SEQ ID NOS: 147
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 31
; LENGTH: 1001
; TYPE: PRT
; ORGANISM: Mammalian (murine/human) SULU3
US-09-291-417-31

Query Match 9.2%; Score 248.5; DB 10; Length 1001;
Best Local Similarity 21.6%; Pred. No. 9.6e-07;
Matches 139; Conservative 117; Mismatches 197; Indels 191; Gaps 29;

Qy 23 PAVAR-----GPGSSQA-PRKPEGAQAARTAGSALRDVSELSRLQLE 64
Db 324 PAVARQEBEEDHGVGRGTGVNSVGSNQSIPMSISASSOSSVNSLPDVSDDKS-BLD 382
Qy 65 DILSTYCVNNQG-----GPGSDG--AOGEPAPBDAKSTYTVARNGEPEPTPVVYGEKE 118
Db 383 MMEGDHTVMSNSSVIHLKPEENYREGDP-----RT---RASDPQSPQVSRHS 430
Qy 119 PSKGDPTTEIRQSDVEGDRHRRPOEKKAKGLGKEIT---LLMQTLNTLSTPEKLA 174
Db 431 HYRNRHFATIRTSLV---TRQMBEHDQSELRQMSGYKMRROHQKQMTLENK 486
Qy 175 ALCKKYA---ELLEHRN---SQQMKLLQKQSQLVQE-----KDLRGHSHK 217
Db 487 AEMDEHRLDKDLETQNNFAEMKLIKKHQAAMEKAKVMSNEKKFOHIAQOQKK 546
Qy 218 AVLARKSLKSLCRLQHRNRSKKEGVOR-----AREBEE---KKEVTSHPV--TLNDI 268
Db 547 EL--NSFLESQREYKLRKEQLKEELNENQSTPKKEQEWLSKQENIQHFOABEENLL 604
Qy 269 QLOME-----QHN-----ERNSKLRQENMELAEKLLRQYELRE 304
Db 605 PRQRYLELCERPKRMMLGRHNLQDLVREELNKRQTKDLEHMLLRQHSQMLEP 664
Qy 305 EHDVKFKHDLQOOLVDAKLQAQAEMLKEAEERHOREKDFLLKEAVESQRMCELMKQOE 364
Db 665 RHLNTIQK---MRCELI---RLQHOTELTNQLEYNKRERELRKHVMEVRRQPKSLKSKE 719
Qy 365 THLKQOL---ALYTEKEEFONTLSKSSSEVFTTFKQEMEKMTKKIKKLEKETTWYSRW 420
Db 720 LQIKKQFODTKIOTQRYKALRNHLE---TTPKSEHKAVLRKLK--EBQTRKLAILA 772
Qy 421 ESSNKALLEMAEBKTVR-----DKELE 442
Db 773 EYDHSINEMLSQALRLDQAEACQVLKMQLOQHELELLNAYQSKIKMQAEACHDREL 832
Qy 443 GLOVKI-----ORLEKLCRALQTERND-----LNKRVODLSAGGGS-----L 480
Db 833 ELEQVSLRRALLLEOKIEEMALQWTERIRSLERQARETEAFDSESRMLGFSNMVL 892
Qy 481 TDSGPERPEPGGAQAPSSPRVTEAPCYGAP--STEASGQTGP 522
Db 893 SNLSPE-----AFSHSYFGASGWSHNPTGGPGP 920

RESULT 33
US-10-177-293-254
; Sequence 254, Application US/10177293
; Publication No. US20030124128A1
; GENERAL INFORMATION:

APPLICANT: Lillie, James
APPLICANT: Glatt, Karen
APPLICANT: Zhao, Xumei
APPLICANT: Gannavarpu, Manjula
APPLICANT: Kamatkar, Shubhangi
APPLICANT: Mertens, Maureen
APPLICANT: Myer, Vic
APPLICANT: Wang, Youzhen
APPLICANT: Xu, Yongyao
APPLICANT: Hoersch, Sebastian
APPLICANT: Monahan, John
APPLICANT: Meyers, Rachel E.
APPLICANT: Bast Jr., Robert C.
APPLICANT: Horobagyi, Gabriel N.
APPLICANT: Pusztai, Lajos
APPLICANT: Meric, Funda
APPLICANT: Sahin, Aysegul
APPLICANT: Mills, Gordon B.
TITLE OF INVENTION: COMPOSITIONS, KITS, AND METHODS FOR IDENTIFICATION, ASSESSMENT,
PREVENTION, AND THERAPY OF BREAST CANCER
FILE REFERENCE: M01-038
CURRENT APPLICATION NUMBER: US/10/177,293
CURRENT FILING DATE: 2002-06-21
PRIOR APPLICATION NUMBER: US 60/299,887
PRIOR FILING DATE: 2001-06-21
PRIOR APPLICATION NUMBER: US 60/301,572
PRIOR FILING DATE: 2001-06-27
PRIOR APPLICATION NUMBER: US 60/306,501
PRIOR FILING DATE: 2001-07-18
PRIOR APPLICATION NUMBER: US 60/325,002
PRIOR FILING DATE: 2001-09-25
PRIOR APPLICATION NUMBER: US 60/362,585
PRIOR FILING DATE: 2002-03-05
PRIOR APPLICATION NUMBER: US 60/xxx,xxx
PRIOR FILING DATE: 2002-05-14
NUMBER OF SEQ ID NOS: 506
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 254
LENGTH: 1005
TYPE: PRT
ORGANISM: Homo sapiens
US-10-177-293-254

Query Match 9.2%; Score 248.5; DB 14; Length 1005;
Best Local Similarity 21.6%; Pred. No. 9.7e-07;
Matches 139; Conservative 117; Mismatches 197; Indels 191; Gaps 29;
Qy 23 PAVEAE-----GPGSQA-PRKPEGQAQTASGALRDVSELSQLE 64
Db 328 PAVEAQEEEEQDHGVRGTGTVNSVGSNQSIPEMSISASSOSSVNSLDPVSDDKS-ELD 386
Qy 65 DILSTYCVNNQ-----GPGEDG--AQGEPAEPEDAESKRTTYVARNGEPEPTPVYGEKE 118
Db 387 MMEGDHTVMSNSSVHLKPEENYREBGP-----RT--RASDPQSPQVSRHKS 434
Qy 119 PSKGDPTTEBROSDEVDGRDHRPQEKKAAGLKEIT-----LLMOTLTLSTPEKLA 174
Db 435 HYRNREHPATIRTSLSV-----TROMQEHEQDSLEQMSGYKMRQHQKQATLENK 490
Qy 175 ALCKKYA-----ELLEHRN--SOKMKLQKQSQVLQV-----KDLRGHSHK 217
Db 491 AEMDEHRLDKDLETQNNFAAEKLIKHOAAMEKAQVMSNEEKKQHQHQAQKK 550
Qy 218 AVLARKSLKSLCRLQRNRSKKEGVOR-----AREEEE---KRKVTSHFQV--TLNDI 268
Db 551 EL--NSFLESQREYKLRKEQLKEELNENQSTPKKEQEWLSKQENIQHFQAEFEANLL 608
Qy 269 QLOME-----QHN-----ERNSKLRQENMELAEKLLKLIQVYELRE 304
Db 609 RRQRYLELCRFRKRMILGRHLEQDLVREELNKRTQKOLEHAMLLRQHESMQLEF 668
Qy 305 EHDVKFKKDLQQLVDKILQQAQEMKAEERHREKDFLLKBAVESQRMCELMKQOE 364
Db 669 RHLNTIQK--MRCELI--RLQHOTELTNQLEYNKRRERELRKHVNEVQPPKSLKSKE 723
Qy 365 THLKQQL-----ALYTEKFEFQNTLSKSEVFTTFKQEMEMTKKIKKLEKETTMYRSRW 420

Db 669 RHLNTIQK--MRCELI--RLQHOTELTNQLEYNKRRERELRKHVNEVQPPKSLKSKE 723
Qy 365 THLKQQL-----ALYTEKFEFQNTLSKSEVFTTFKQEMEMTKKIKKLEKETTMYRSRW 420
Db 724 LQIKKQFQDTCKIQTRQYKALRNHLE-----TTPSEKHAVLKRLK--EEQTRKLAILA 776
Qy 421 ESSNKALLEKABEKTVR-----DKELE 442
Db 777 EYDHSINEMLSQALRLDEAQBACQVLKMQLOQSELELLNAYQSKIKMQAQAQHDRELR 836
Qy 443 GLOVKI-----QRLEKLCRALQTERND-----LNKRVQDLSAGQGS-----L 480
Db 837 ELQORVSLRALLEQKLEEMALQNERTERNSLLERQARIEAEFDSMRGLGFSNNWL 896
Qy 481 TDSGPRRPRPGCAQAPSPRVTEAPCYGAP--STEASGQGTGP 522
Db 897 SNLSP-----AFSHSYFGASGWSHNP7GGPG 924
RESULT 34
US-10-317-835-18
; Sequence 18; Application US/10317835
; Publication No. US20030157531A1
; GENERAL INFORMATION:
; APPLICANT: Exelixis, Inc.
; TITLE OF INVENTION: TAOJIKS AS MODIFIERS OF THE BETA-CATENIN PATHWAY AND METHODS OF
; FILE REFERENCE: EX02-141C
; CURRENT APPLICATION NUMBER: US/10/317,835
; CURRENT FILING DATE: 2002-12-12
; PRIOR APPLICATION NUMBER: 60/340,312
; PRIOR FILING DATE: 2001-12-13
; NUMBER OF SEQ ID NOS: 18
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 18
; LENGTH: 1005
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-317-835-18

Query Match 9.2%; Score 248.5; DB 14; Length 1005;
Best Local Similarity 21.6%; Pred. No. 9.7e-07;
Matches 139; Conservative 117; Mismatches 197; Indels 191; Gaps 29;

Qy 23 PAVEAE-----GPGSQA-PRKPEGQAQTASGALRDVSELSQLE 64
Db 328 PAVEAQEEEEQDHGVRGTGTVNSVGSNQSIPEMSISASSOSSVNSLDPVSDDKS-ELD 386
Qy 65 DILSTYCVNNQ-----GPGEDG--AQGEPAEPEDAESKRTTYVARNGEPEPTPVYGEKE 118
Db 387 MMEGDHTVMSNSSVHLKPEENYREBGP-----RT--RASDPQSPQVSRHKS 434
Qy 119 PSKGDPTTEBROSDEVDGRDHRPQEKKAAGLKEIT-----LLMOTLTLSTPEKLA 174
Db 435 HYRNREHPATIRTSLSV-----TROMQEHEQDSLEQMSGYKMRQHQKQATLENK 490
Qy 175 ALCKKYA-----ELLEHRN--SOKMKLQKQSQVLQV-----KDLRGHSHK 217
Db 491 AEMDEHRLDKDLETQNNFAAEKLIKHOAAMEKAQVMSNEEKKQHQHQAQKK 550
Qy 218 AVLARKSLKSLCRLQRNRSKKEGVOR-----AREEEE---KRKVTSHFQV--TLNDI 268
Db 551 EL--NSFLESQREYKLRKEQLKEELNENQSTPKKEQEWLSKQENIQHFQAEFEANLL 608
Qy 269 QLOME-----QHN-----ERNSKLRQENMELAEKLLKLIQVYELRE 304
Db 609 RRQRYLELCRFRKRMILGRHLEQDLVREELNKRTQKOLEHAMLLRQHESMQLEF 668
Qy 305 EHDVKFKKDLQQLVDKILQQAQEMKAEERHREKDFLLKBAVESQRMCELMKQOE 364
Db 669 RHLNTIQK--MRCELI--RLQHOTELTNQLEYNKRRERELRKHVNEVQPPKSLKSKE 723
Qy 365 THLKQQL-----ALYTEKFEFQNTLSKSEVFTTFKQEMEMTKKIKKLEKETTMYRSRW 420

Db 724 LOIKKQFQDCKIQTRQYKALRNHLE-----TTPKSEHKAVALKRLK--BEQTRKLAIAL 776
Qy 421 ESSNKALLEMAEETVR-----DKELE 442
Db 777 EYDHSINEMLSQALRLDEAQAECQVLMQQLQOELELNAVCSKIKMQAQAQHDREL 836
Qy 443 GLQVKI-----QRLKLCRALQTERND-----LNKRVODLSAGQGS-----L 480
Db 837 ELEQVSLRALLQKTEEMMALQNERIRISLLERQARETEAFDSESMRLGFSNMVL 896
Qy 481 TDSGPRRPEGQAQAFSSPRVTEAPCPGAP--STEASGTGP 522
Db 897 SNLSPE-----APSHSYPCASGWSHNPCTGPGP 924

RESULT 35
US-10-049-3413
; Sequence 3413, Application US/10264049
; Publication No. US20040005579A1
; GENERAL INFORMATION:
; APPLICANT: Birse et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: PA133P1
; CURRENT APPLICATION NUMBER: US/10/264,049
; CURRENT FILING DATE: 2002-10-04
; PRIOR APPLICATION NUMBER: PCT/US01/18569
; PRIOR FILING DATE: 2001-06-07
; PRIOR APPLICATION NUMBER: US 60/209,467
; PRIOR FILING DATE: 2000-06-07
; NUMBER OF SEQ ID NOS: 4360
; SOFTWARE: Patent In Ver. 3.1
; SEQ ID NO 3413
; LENGTH: 85
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE: MISC_FEATURE
; NAME/KEY: MISC_FEATURE
; LOCATION: (1)
; OTHER INFORMATION: Xaa equals any of the twenty naturally occurring L-amino acids
; FEATURE: MISC_FEATURE
; NAME/KEY: MISC_FEATURE
; LOCATION: (3)
; OTHER INFORMATION: Xaa equals any of the twenty naturally occurring L-amino acids
; FEATURE: MISC_FEATURE
; NAME/KEY: MISC_FEATURE
; LOCATION: (18)
; OTHER INFORMATION: Xaa equals any of the twenty naturally occurring L-amino acids
; FEATURE: MISC_FEATURE
; NAME/KEY: MISC_FEATURE
; LOCATION: (75)
; OTHER INFORMATION: Xaa equals any of the twenty naturally occurring L-amino acids
US-10-049-3413

Query Match 9.1%; Score 247; DB 15; Length 85;
Best Local Similarity 64.5%; Pred. No. 5.8e-08;
Matches 49; Conservative 13; Mismatches 14; Indels 0; Gaps 0;
Qy 347 LKAVESQRCMELMKQOETHLKKQALALYKTEFEQNTLSKSEVFTTFQEMKMTKKI 406
Db 2 LKAVESQRCMELMKQOETHLKKQALALYKTEFEQNTLSKSEVFTTFQEMKMTKKI 61
Qy 407 KLEKETTMYRSWES 422
Db 62 KLEKETTMYRSWES 77

RESULT 36
US-10-445-735-2
; Sequence 2, Application US/10445735
; Publication No. US20040009175A1
; GENERAL INFORMATION:
; APPLICANT: Cobb, Melanie
; TITLE OF INVENTION: TAO PROTEIN KINASE POLYPEPTIDES AND METHODS OF USE THEREOF

; FILE REFERENCE: 10624-026-999
; CURRENT APPLICATION NUMBER: US/10/445,735
; CURRENT FILING DATE: 2003-05-27
; PRIOR APPLICATION NUMBER: US/09/686,346
; PRIOR FILING DATE: 2000-10-10
; PRIOR APPLICATION NUMBER: 09/060,410
; PRIOR FILING DATE: 1998-04-14
; NUMBER OF SEQ ID NOS: 26
; SOFTWARE: Patent In version 3.0
; SEQ ID NO 2
; LENGTH: 1001
; TYPE: PRT
; ORGANISM: Rattus norvegicus
; US-10-445-735-2

Query Match 9.1%; Score 247; DB 15; Length 1001;
Best Local Similarity 21.9%; Pred. No. 1.2e-06;
Matches 142; Conservative 118; Mismatches 212; Indels 176; Gaps 30;
Qy 23 PAVEAE-----GPGSSQA-PRKEGQAQARTAQSGALRDVSEELSRLE 64
Db 324 PAVEAQEEREEODHGGGRTGTVNSVGSNQSIPSMISASSQSSSVNSLPDASDDKS-ELD 382
Qy 65 DILSTYCVDNNG-----GPGEDGAQCEPAEPEDAEKSRITYVARNGEPEPTPVVYGEKPS 120
Db 383 MMEGDHTVMSNSSVHLKPEENYQ-EGGDP-----TRASAPQSPQVSRHSHY 432
Qy 121 KQDPNTEIRQSDVGDHRRPQEKKAKGLKEIT-----LLMQTLNTLSPEKLAAL 176
Db 433 RNREHFATIRASLV-----TROMQEHQDSELRQMSGYKMRQRHQKQMTLENKLKAE 488
Qy 177 CKKYA-----ELLEERN--SOKOMLKQKQQLVQE-----KDLRGEHSAV 219
Db 489 MDEHRLDLXOLETORNNFAEMEKLIKHOASMEKAKVMAEKKFOHQAQOQKEL 548
Qy 220 LARSKLSLCRELQRHNSLKEGVOR-----AREEB-----KKVTSVSHQV--TLNDIOL 270
Db 549 --NSFLESQREYKLRKEQLKEELNENOSTPKKEQEWLSKQENI QHFQAEENALLR 606
Qy 271 QME-----OHN-----BRNSKLQENMELAEKRLKLIQOYELREH 306
Db 607 QRYLELCRFRFRMLGRHMLQDLVREELNKQTKQDLEHMLLQGHESMQELEPH 666
Qy 307 IDKVPFKHDLQOQLVDKILQQAQEMLEKAEERHOREKDFLLKAEVSEQRCMELMKQOETH 366
Db 667 LNTIQK--MRCELI--RLQHQTELNTQLEYNKREERELRRKHVMVEVRQOQPSLAKSEIQ 721
Qy 367 LKQOL-----ALYKTEFEFEQNTLSKSEVFTTFQEMKMTKKIKLEKETTMYRSWES 422
Db 722 IKQFQDCKIQTRQYKALRNHLE-----TTPKSEHKAVALKRLK--BEQTRKLAIALQ 774
Qy 423 SNKALLEMAEETVR-----DKELEGL 444
Db 775 YDHSINEMLSQALRLDEAQAECQVLMQQLQOELELNAVCSKIKMQAQAQHDREL 834
Qy 445 QVKI-----QRLKLCRALQTERND-----LNKRVODLSAGQGS-----LTD 482
Db 835 EQRVSLRALLQKTEEMMALQNERIRISLLERQARETEAFDSESMRLGFSNMVL 894
Qy 483 SGPRRPEG-PCAQAPS--SPRVTEAP--CYGAPSTEASG---QTCGP 523
Db 895 LSPEAFSHSYPCASGWSHNPCTGSGPHNGHPMGTPQAWHPMQGPGP 942

RESULT 37
US-09-925-299-988
; Sequence 988, Application US/09925299
; Patent No. US20020055627A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins and Antibodies
; FILE REFERENCE: PA102
; CURRENT APPLICATION NUMBER: US/09/925,299

466 SOVESKQNAELAKLQOE--LSKSVKELVEKSAVRQDEQORCALEKAAAFKQVLOQAS 524
339 HOREKDFLLKEAVESQRMCELMKQOETHLKOOLALYTEKFEFFONTLS-----KSSEVP 392
525 HRESEALQKRLDEVR--ELCHTQSSH--ASLRADAERKAOEQOQOQMAELHSLKQSSAE 580
393 TTFK-----QEMEKNTKKIKLEKETTMYRSRWESSNKALLEMAEKTVR 437
581 VRSKCEBSGLHGQLOEBAENSQUTERIRSEALLEAGARDADQDVQASQXEAQOQTR 640
438 DKELE---GLQVKIQRLKLCRALQTERNDLN-----KRVQDLSAGGQ-----SLTDS 483
641 LKELESQVSGLEKEALELREAVBQKVNDLRKNWKAEMALATAEQACKELHSLTQA 700
484 GPERRP-----EGPCAQA-----PSSPRVTEAPCYGAPST 514
701 KESEKQICLLEAQTMEALLALLPELSVLAQONYTEWLODLKEKGTLLKHPAPAPBSS 760
515 BASQGTGPQEPSTA 528
761 DLASKLREAEETQS 774

RESULT 38
US-09-925-299-988
; Sequence 988, Application US/09925299
; Publication No. US20030040617A9
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins and Antibodies
; FILE REFERENCE: PA102
; CURRENT APPLICATION NUMBER: US/09/925,299
; CURRENT FILING DATE: 2001-08-10
; PRIOR APPLICATION NUMBER: PCT/US00/05883
; PRIOR FILING DATE: 2000-03-08
; PRIOR APPLICATION NUMBER: 60/124,270
; PRIOR FILING DATE: 1999-03-12
; NUMBER OF SEQ ID NOS: 1556
; SOFTWARE: Patent In Ver. 2.0
; SEQ ID NO 988
; LENGTH: 909
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURES:
; NAME/KEY: SITE
; LOCATION: (32)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
; NAME/KEY: SITE
; LOCATION: (41)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
; NAME/KEY: SITE
; LOCATION: (47)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
; NAME/KEY: SITE
; LOCATION: (48)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
; NAME/KEY: SITE
; LOCATION: (52)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
; NAME/KEY: SITE
; LOCATION: (58)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
; NAME/KEY: SITE
; LOCATION: (62)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
; NAME/KEY: SITE
; LOCATION: (125)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
; NAME/KEY: SITE
; LOCATION: (632)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
; NAME/KEY: SITE

CURRENT FILING DATE: 2001-03-10
PRIOR APPLICATION NUMBER: PCT/US00/05883
PRIOR FILING DATE: 2000-03-08
PRIOR APPLICATION NUMBER: 60/124,270
PRIOR FILING DATE: 1999-03-12
NUMBER OF SEQ ID NOS: 1556
SOFTWARE: Patent In Ver. 2.0
SEQ ID NO 988
LENGTH: 909
TYPE: PRT
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: SITE
LOCATION: (32)
OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
NAME/KEY: SITE
LOCATION: (41)
OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
NAME/KEY: SITE
LOCATION: (47)
OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
NAME/KEY: SITE
LOCATION: (48)
OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
NAME/KEY: SITE
LOCATION: (52)
OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
NAME/KEY: SITE
LOCATION: (58)
OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
NAME/KEY: SITE
LOCATION: (62)
OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
NAME/KEY: SITE
LOCATION: (125)
OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
NAME/KEY: SITE
LOCATION: (632)
OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
NAME/KEY: SITE
LOCATION: (851)
OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
US-09-925-299-988

Query Match 9.0%; Score 244.5; DB 9; Length 909;
Best Local Similarity 22.4%; Pred. No. 1.5e-06;
Matches 151; Conservative 109; Mismatches 237; Indels 177; Gaps 32;

1 KSSPGQPEAG--PEGQAEPSQAAPAVE--AEGPGSSQAPRPEGAQTAQ-----48
132 KAEGAGNQKKAEGAGNQKKAEGAGNQKQKGEAGNQKKGTEGAQKKAERSPNQKK 191
49 -----SGALRDVSELSRQLEDILSTYCDVNNQGGFGBDGAQGEPAEPEDAE-----K 96
192 GEGAPIQKKKADSVANQGTKEGI-----TNQG-----KKAEGSPSEGGKKAEGSPNQKK 240
97 SRTYVARNPEPTPVYVYGEKESKGDPTNTEBETROSDEVDGDRDHRPEQKK---KAKG---151
241 KADAAANQKKTESASVQGR-----NT-DVAQSPFAPKQOE--APAKKSGSKKKGP 289
152 -----LGEKITLLAQTLNTLSTPE---EKLAAICKKYVAILER--HRNSQK--QMKL 196
290 GPPDADGPLYLPYKTLVSTGVSVFNEGEAQLIEILSEKAGIIOQTHKATQKGPVAI 349
197 LQKKQSQVQKDLRGEHSKAVLARKSLKSLCREL--QRHRSLEKRGVOR---ARESE 251
350 LKR---QLEEKELLATQEDADAVAKSKLRELKEMAAEKAKAAAGAEAKVKQLVAREGE 406
252 -----EKEKEVTSHPQVTLNDIQLQMEQ--HNERNSKLQRNEMELAEKLI 297
407 ITAVQRMQASREHVKEV--QOLQKIRVLOEQLNGPNTQLARLQOENSILDALNQAT 465
298 EQYE-----LREHIDKVPK-----HFDLQOQLVDKLLQQAQEMLEKEAER 338

LOCATION: (851)
OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
US-09-925-299-988

Query Match 9.0%; Score 244.5; DB 10; Length 909;
Best Local Similarity 22.4%; Pred. No. 1.5e-06;
Matches 151; Conservative 109; Mismatches 237; Indels 177; Gaps 32;

QY 1 KSSPGQPEAG--PEGAGPSPQAPAVE--AEGPGSSQAPRPEGAQARTAQ-----48
DB 132 KXAGGQNGQKAGGQNGQKAGGQNGQKAGGQNGQKAGGQNGQKAGGQNGQKAGGQNGQK 191
QY 49 -----SGALRVSBELSRQLEILSTYCVNNQGGPBGDGAOGEPAPEDAR-----K 96
DB 192 GEGAPIQKADSVANGQTKVEGI-----TNQG-----KXAGSPSEKKAAGSPNOGK 240
QY 97 SRYTVARNGEPEPTPVYVYGESEKSGDPNTEBTRQSDVGDHRRPEQKK---KAKG--151
DB 241 KADAAANQKATSEASVQGR-----NT-DVAQSPAPKQZ--APAKKSGKKKGEP 289
QY 152 -----LQKETTLLMOTLNTLSTPE---EKLAALCKKYAELEL--HRNSQK--OMKL 196
DB 290 GPPDADGPLYLYPYKTLVSTVGSVMVFNESGAEQRLIEILSEKAGLIQDTHKATQKGDPAI 349
QY 197 LQKQSQLVQEKDHLRGHSHKAVLARSKLESICREL--QHNRSLEEGVOR---AREEB 251
DB 350 LXR-----QLESEKELLATQEDDAVAKSKURELNKEMAAEKAKAAAGAEKVYKQLVAREQB 406
QY 252 -----EKREVTSHFQVTLNDIQLQMEQ--HNERNSKLQENNELAERLKLI 297
DB 407 ITAVQAEQASVREHYKEV--QQLQKIRTIQEQLENGPNTQLARLQENSIIRDALNQAT 465
QY 298 EQVE-----LREHIDKVKF-----HKDLOQLVDKLOQAQEMLEKAEER 338
DB 466 SOVESKQNAELAKLRQ--LSKVSKELVKSEAVRQDEQKKALEAKAAAFKQVLQAS 524
QY 339 HOREKDFLLKAEVSRQMCMLKQETHLQKQALALYKEPEPNTLS-----KSEVP 392
DB 525 HRSEELQKLEDEVR--ELCHTQSH--ASLRADAERKAEQEQQVABLSKLSQSEAE 580
QY 393 TTFK-----QEMEMTKIKKLEKETTMTSRWESSNKALLENAEKTVR 437
DB 581 VRSKCEBLSGLHGLQAPARAENSQLTIRISIEALLEAGQARDADQVQASXEAQQOQTR 640
QY 438 DKELE-----GLOVKIQRLKALOTERNDLN---KRVODLSAGQGG-----SLTDS 483
DB 641 LKELESQVSGLEKAEIELREAVPQKVKYNDLREKQWKAMEALATAEQACKEXLHSLTQA 700
QY 484 GPERRP-----EGPGAQA-----PSSPRVTEAPCYGAPST 514
DB 701 KESEKQLCLTEAGTMEALLALLPELSVLAQNVTEWLODLKXGPTLLKHPAPAPSPS 760
QY 515 EASQQTGPQETSA 528
DB 761 DLASKLREABETQS 774

RESULT 39
US-10-342-136-1

Sequence 1, Application US/10342136
Publication No. US20030199092A1
GENERAL INFORMATION:
APPLICANT: Meyer, David I.
TITLE OF INVENTION: METHOD FOR INCREASING MRNA HALF-LIFE IN EUKARYOTIC CELLS
FILE REFERENCE: 407T-301610US
CURRENT APPLICATION NUMBER: US/10/342,136
CURRENT FILING DATE: 2003-01-13
PRIOR FILING DATE: 2002-01-11
NUMBER OF SEQ ID NOS: 3
SOFTWARE: PatentIn version 3.2
SEQ ID NO 1
LENGTH: 1534

TYPE: PRT
ORGANISM: Canis familiaris
US-10-342-136-1

Query Match 9.0%; Score 242; DB 14; Length 1534;
Best Local Similarity 20.4%; Pred. No. 4e-06;
Matches 153; Conservative 114; Mismatches 222; Indels 260; Gaps 30;

QY 1 KSSPGQPEAGPE-----GAQERPSQAPAVEAGPGSSQAP 36
DB 710 KXSEGPSNGKVKVDASANQSKRAESAPITQGNADVMVQSEAPKQEPAPAKTKSGSKGGE 769
QY 37 RKP-----EGAAQAR-----TAQSCALRDVSEB-----58
DB 770 GPPDSDSPLYLYPYKTLVSTVGSVMVFNESGAEQRLIEILSEKAGVIQDTHKATQKGDPAI 829
QY 59 LSRQLED---ILSTYCVNNQGGPBGDGAOGEPAPEDAKSRT---YVARNGEPEPTPV 112
DB 830 LKRLQLEEKLLAT-----EQEDAAVAKSKLREVNKELAAKAKA 869
QY 113 VYGEKEPSK---GDPNTEIROSDEVDHRRPQEKKAK-----G 151
DB 870 AAGEAKVYKQLVAREQETAVQARTEASYREHVKEVQQLQKIRTIQEQLENGPNTQLAR 929
QY 152 LQKETTLLMOTLN-----TLSTPEKLAALCKKYAELEHNRNSQKMKLLQK 199
DB 930 LQENSIIRDALNQATSOVESKQNTLAKRLQELSKVSEKELVEKSEAAARQEQKALLET 989
QY 200 KOSQVLQKDHRLRGHSHKAVLA-RSKLESICREL---QHNRSLEKEGVORARBEKEEK 255
DB 990 KTAALQKQVLQASHKESSEALQKLDSEVRSQTSQSHASLRAD-AEKAQEQEQQQA 1048
QY 256 EYTSHPQVTLNDIQLQMEQHNRNSKL---ROENNELAERLKK---LIEQVELRE-----304
DB 1049 ELHSLKLSSEAEVSKSELSLHQLKAEARNSQLMERIRSIEALLEAGQARDTQAO 1108
QY 305 -----EHIDKVKFKDLOQLVDKLOQAQEMLEKAEERHQ-----REKDFLLKAEVES- 353
DB 1109 ASRAEQHQAQL---KELSEQV--WCLEKEATELKEAVEQKVKYNDLREKNWKAMEALASA 1163
QY 354 QMCILMKQETHLQKQ-----LAL-----YTE-----376
DB 1164 ERACEKLSLTQAKSESEKQLSLTEAQTKKALLALLPALSSAPQSYTEWLQELREKGP 1223
QY 377 -----KFEFQNTLSKSEVFTTFKQEMKMTKIKK-LEKETT 414
DB 1224 ELLKQRPADTDPSSDLASKLREABETQNNLOAECDOVRTILAEETGLMLDLQKSYEEBQ 1283
QY 415 MYRSWESSNKALLEMAEKTVRDKLEGLQVKIQ-----RLEKLCR 456
DB 1284 V-----WKAVSATEBELOKSRVTYVHLSDIVKELGELSESQVREHSHLEAELEKMA 1339
QY 457 ALQTERNDLNKRVQDLSAGQGGSLTDSGPERPPEGPGAPSS-----PRVTEAPCY--508
DB 1340 AASACQSVAKV-----AGLRQLLLESQSLDAAKSEAKQSQNELALVRQLSKSHVE 1395
QY 509 -----PGAPSTHAGQTPGPQETSA 530
DB 1396 DGDVAGSPAAPPAAE-----QDPVELKA 1417

RESULT 40

US-10-205-219-102
Sequence 102, Application US/10205219
Publication No. US20030138803A1
GENERAL INFORMATION:
APPLICANT: Warner-Lambert Company
APPLICANT: Lee, Kevin
APPLICANT: Dixon, Alistair
APPLICANT: Brooksbank, Robert
APPLICANT: Finnock, Robert
TITLE OF INVENTION: Identification and Use of Molecules Implicated in Pain
FILE REFERENCE: WL-A-018200

```
; CURRENT APPLICATION NUMBER: US/10/205,219
; CURRENT FILING DATE: 2002-07-24
; PRIOR APPLICATION NUMBER: GB 0118354.0
; PRIOR FILING DATE: 2001-07-27
; NUMBER OF SEQ ID NOS: 197
; SOFTWARE: Patent In Ver. 2.1
; SEQ ID NO 102
; LENGTH: 892
; TYPE: PRT
; ORGANISM: Rattus norvegicus
; FEATURE:
; OTHER INFORMATION: Myosin heavy chain
US-10-205-219-102

Query Match
Best Local Similarity 8.8%; Score 238; DB 14; Length 892;
Matches 113; Conservative 88; Mismatches 159; Indels 164; Gaps 18;

QY 33 SQAPKPGCAQRTAQSGALRDVSEHLSRQLEDILSTYCVNNOGGPGGCGAGGPAEPE 92
DB 45 SNLQSKIEDQALGNQ---LQKIKELQARIEEL-----EEIEASRASRA 87
QY 93 DAESRTYVANGSGEPTPVVYGEKPSKGDENTEEIRQS--DEVGDRDRHRRPOEKKKAG 151
DB 88 KAEKQSDLSR-----ELEEISERLEAGGATSQIEMWKKREA 126
QY 152 LKKEITLLMOTINTLSTPEKLAALCKK-----VAELLEHRNSQOMKLLQKQSQVLQ 206
DB 127 ---EFQKRRDRLEATLQHEATAATLAKKHAQSVAGELGEQIDNLRVQKLEKEKSEMGM 183
QY 207 EKDHGRGSHKAVLARSK--LESCLRLQ-----233
DB 184 EIDDLAS--NMEVLSKSGNLEKMCRTLEDOVSELKTEBEQQRLINELTAQRGLQTES 241
QY 234 -RHNSLKE--EGVQARBEERKEKVTSH-FQVTLNDIQLQMFQHNERNKSLQENNELA 272
DB 242 GEYSRQLDEKDSVLSQSRGKQAFQTOQIEELKRLQLEBEVKAASALAHASSHDCDLR 301
QY 273 ECHNER-----NSKLR-----QENNELAERLKKLEQVELBEHI 307
DB 302 EYBEEQAEKAELOFAMSKANSEVAQWTKYETDAIQSTELBEAKKLAQRLQDAEHHV 361
QY 308 DKVPKHDLQOOLVDKALQQAEMLKAEERHQREKDFLKEAVESQRMCELMKQOETHL 367
DB 362 -----EAVNAKAS-----LEKTKORLQNEVEDLMIDVERTWAACAALDKQRNF 406
QY 368 KQALALYTEKPEEFQNTLSKSSEVFTTFKQEMERKTKIKKLEKTTMYRSWESSNKAL 427
DB 407 DKLLAEWKQYBETH-----AELEASQKESRSISTELFKINAYERSILDQ 452
QY 428 LEMAEKTVRQKEGL--QV-----KIQRLKLCRALQTERNDLNKRVQDLQA 474
DB 453 ETLKRENKNOQEI--SDLETEQIABGGKRIHELEKIKQIQEOKSLEQALBEAEA 506

RESULT 41
US-10-408-765A-1168
; Sequence 1168, Application US/10408765A
; Publication NO. US20040101874A1
; GENERAL INFORMATION:
; APPLICANT: Ghosh, Soumitra S.
; APPLICANT: Fahy, Robin D.
; APPLICANT: Zhang, Bing
; APPLICANT: Gibson, Bradford W.
; APPLICANT: Taylor, Steven W.
; APPLICANT: Glenn, Gary M.
; TITLE OF INVENTION: TARGETS FOR THERAPEUTIC INTERVENTION
; TITLE OF INVENTION: IDENTIFIED IN THE MITOCHONDRIAL PROTEOME
; FILE REFERENCE: 660088.465
; CURRENT APPLICATION NUMBER: US/10/408,765A
; CURRENT FILING DATE: 2003-04-04
; NUMBER OF SEQ ID NOS: 3077

; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 1168
; LENGTH: 1938
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-408-765A-1168

Query Match
Best Local Similarity 8.8%; Score 237; DB 16; Length 1938;
Matches 111; Conservative 113; Mismatches 197; Indels 98; Gaps 20;

QY 19 SQAPPAVEAGGSGSQAPKPGCAQRTAQSGALRDVSE-----LSR 61
DB 1092 SOLQARIDDEQVHSLQFQKIKELQAR-----FEELERTEAEHTLRAKIEKQSDLAR 1145
QY 62 QLEDILSTYCVNNOGGPGGCGAGGPAEPEDAE--KSRTYVARNGEPEPTPVVYGEKE 118
DB 1146 ELSEI-----SERLEBASGATSQIEMWKKREAFQWRDRLEATLQHEATAATLRKKQ 1200
QY 119 PSKGDPTTEIRQSDVGDHRRHRRPOEKKKAG--LGKEITLLMOTINTLSTPEKLAALC 177
DB 1201 -----ADSVAGELGEQIDNLRVQKLEKEKSELKMEIDDMASNIKSKSNIBRTC 1253
QY 178 KYIAELLEHEHNSQOMKLLQKQSQVLV---OEKHLRGEHSHKAVLARSKLSLCLRELQ 233
DB 1254 RTVEDQPE-----IAKDEQSTQLHDLNMOKARLOTQNGELSHRVEKESLSLSQLT 1306
QY 234 RHNSLKE--EGVQARBEERKEKVTSH-FQVTLNDIQLQMFQHNERNKSLQENNELA 290
DB 1307 KSKQALQQLBELKQWEEETKAKNAHAHALQSSRHDCDLIREQYBEE---QEAQALQ 1362
QY 291 ERLKKL---IQEYLRERHHDVFKHDLQO--OLVDAKQQAOS-----MLKEA 335
DB 1363 RALSKANSEVAQWTKYE--TDAIQSTELBEAKKLAQRLQDAEETETANSKASLEKT 1421
QY 336 BERHQREKDFLKEAVESQRMCELMKQOETHLKOOLALYTEKFEFPQNTL-----SK 387
DB 1422 KQRLQGEVEDLNRDLERSHTACATLDKKQRNFDKVLAEWKQKLDQESQAELEAAQKESRSL 1481
QY 388 SSEYF---TFKQ--SMERKTKIKKLEKTTMYRSWESSNKALLEMAEKTVRDKEL 441
DB 1482 STELFKRNAYREVVDQLETLRRENKLOBEISDLTEQIATGKQLQAEAKTKLVEQEK 1541
QY 442 EQLQVKIQRLE-----KLQRA---IQTERNDLNKRV 469
DB 1542 SDLQVALEEVGSLHEBESKILRVQELSQVKSLEDRKV 1580

RESULT 42
US-10-336-472-16
; Sequence 16, Application US/10336472
; Publication NO. US20040043929A1
; GENERAL INFORMATION:
; APPLICANT: Anderson, David W.
; APPLICANT: Baumbgartner, Robert A.
; APPLICANT: Burgess, Catherine E.
; APPLICANT: Casman, Stacie J.
; APPLICANT: Chant, John S.
; APPLICANT: Bergths, Constance
; APPLICANT: Gangolli, Esha A.
; APPLICANT: Edinger, Shlomit R.
; APPLICANT: Ellerman, Karen
; APPLICANT: Furtak, Katarzyna
; APPLICANT: Gerlach, Valerie
; APPLICANT: Gilbert, Jennifer A.
; APPLICANT: Gunther, Erik
; APPLICANT: Gorman, Linda
; APPLICANT: Guo, Xiaojia Sasha
; APPLICANT: Ji, Weizhen
; APPLICANT: Li, Li
; APPLICANT: Liu, Xiaohong
; APPLICANT: Miller, Charles B.
```

; Sequence 6752, Application US/09543681A
; Patent No. 6605709
; GENERAL INFORMATION:
; APPLICANT: GARY BRETON
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PROTEUS MIRABILIS
; FILE REFERENCE: 2709.1002-001
; CURRENT APPLICATION NUMBER: US/09/543.681A
; CURRENT FILING DATE: 2000-04-05
; PRIOR APPLICATION NUMBER: US 60/128,706
; PRIOR FILING DATE: 1999-04-09
; NUMBER OF SEQ ID NOS: 8344
; SEQ ID NO 6752
; LENGTH: 474
; TYPE: PRT
; ORGANISM: Proteus mirabilis
US-09-543-681A-6752

Query Match 1.3%; Score 7; DB 4; Length 474;
Best Local Similarity 100.0%; Pred. No. 3.6e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 128 BIRQSD 134
Db 386 BIRQSD 392

RESULT 47
US-08-962-859A-2
; Sequence 2, Application US/08962859A
; Patent No. 6127345
; GENERAL INFORMATION:
; APPLICANT: Burnham, Martin K. R.
; TITLE OF INVENTION: No. 6127345el Glucose 6-Phosphate
; FILE REFERENCE: Dehydrogenase Gene
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Dechert, Price & Rhoads
; STREET: 4000 Bell Atlantic Tower, 1717 Arch Stre
; CITY: Philadelphia
; STATE: PA
; COUNTRY: USA
; ZIP: 19103-2793
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: Windows
; SOFTWARE: FastSeq for Windows Version 2.0b
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/962,859A
; FILING DATE: 03-NOV-1997
; CLASSIFICATION: 514
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 60/035,072
; FILING DATE: 21-JAN-1997
; ATTORNEY/AGENT INFORMATION:
; NAME: Jackson, Arthur E
; REGISTRATION NUMBER: 34,354
; REFERENCE/DOCKET NUMBER: GM50001
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 609/520-3254
; TELEFAX: 609/520-3259
; TELEX:
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 495 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
US-08-962-859A-2

Query Match 1.3%; Score 7; DB 3; Length 495;
Best Local Similarity 100.0%; Pred. No. 3.8e+02;

Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 48 QSGALD 54
Db 228 QSGALD 234

RESULT 48
US-09-252-991A-20533
; Sequence 20533, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; FILE REFERENCE: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 20533
; LENGTH: 526
; TYPE: PRT
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-20533

Query Match 1.3%; Score 7; DB 4; Length 526;
Best Local Similarity 100.0%; Pred. No. 4e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 181 AELLEEH 187
Db 122 AELLEEH 128

RESULT 49
US-08-285-440-5
; Sequence 5, Application US/08285440
; Patent No. 5532337
; GENERAL INFORMATION:
; APPLICANT: Ken'ichiro HAYASHI et al.
; TITLE OF INVENTION: POLYPEPTIDE
; NUMBER OF SEQUENCES: 32
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Wenderoth, Lind & Ponack
; STREET: 805 Fifteenth Street, N.W., #700
; CITY: Washington
; STATE: D.C.
; COUNTRY: U.S.A.
; ZIP: 20005
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 5.25 inch, 500 kb
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: MS-DOS
; SOFTWARE: DisplayWrite
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/285,440
; FILING DATE:
; CLASSIFICATION: 536
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 07/858,947
; FILING DATE: March 27, 1992
; ATTORNEY/AGENT INFORMATION:
; NAME: Warren M. Cheek, Jr.
; REGISTRATION NUMBER: 33,367
; REFERENCE/DOCKET NUMBER:
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 202-371-8850
; TELEFAX:
; TELEX:

INFORMATION FOR SEQ ID NO: 5;
SEQUENCE CHARACTERISTICS:
LENGTH: 532 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: peptide
HYPOTHETICAL:
ANTI-SENSE:
FRAGMENT TYPE:
ORIGINAL SOURCE:
ORGANISM:
STRAIN:
INDIVIDUAL ISOLATE:
DEVELOPMENTAL STAGE:
HAPLOTYPE:
TISSUE TYPE:
CELL TYPE:
CELL LINE:
ORGANELLE:
IMMEDIATE SOURCE:
LIBRARY:
CLONE:
POSITION IN GENOME:
CHROMOSOME/SEGMENT:
MAP POSITION:
UNITS:
FEATURE:
NAME/KEY:
LOCATION:
IDENTIFICATION METHOD:
OTHER INFORMATION:
PUBLICATION INFORMATION:
AUTHORS:
TITLE:
JOURNAL:
VOLUME:
ISSUE:
PAGES:
DATE:
DOCUMENT NUMBER:
FILING DATE:
PUBLICATION DATE:
RELEVANT RESIDUES IN SEQ ID NO:
US-08-285-440-5

Query Match 1.3%; Score 7; DB 1; Length 532;
Best Local Similarity 100.0%; Pred. No. 4e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 248 REBERK 254
Db 335 REBERK 341

RESULT 50
US-08-630-349-5
Sequence 5, Application US/08630349
Patent No. 5739008
GENERAL INFORMATION:
APPLICANT: Ken'ichiro HAYASHI et al.
TITLE OF INVENTION: POLYPEPTIDE
NUMBER OF SEQUENCES: 32
CORRESPONDENCE ADDRESS:
ADDRESSEE: Wenderoth, Lind & Ponack
STREET: 805 Fifteenth Street, N.W., #700
CITY: Washington
STATE: D.C.
COUNTRY: U.S.A.
ZIP: 20005
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette, 3.5 inch, 144 mb
COMPUTER: IBM Compatible

OPERATING SYSTEM: MS-DOS
SOFTWARE: Wordperfect 5.1
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/630,349
FILING DATE: April 10, 1996
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/285,440
FILING DATE: August 4, 1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 07/858,947
FILING DATE: March 27, 1992
ATTORNEY/AGENT INFORMATION:
NAME: Warren M. Cheek, Jr.
REGISTRATION NUMBER: 33,367
INFORMATION FOR SEQ ID NO: 5:
SEQUENCE CHARACTERISTICS:
LENGTH: 532 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: peptide
US-08-630-349-5

Query Match 1.3%; Score 7; DB 1; Length 532;
Best Local Similarity 100.0%; Pred. No. 4e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 248 REBERK 254
Db 335 REBERK 341

RESULT 51
US-09-138-277C-3
Sequence 3, Application US/09138277C
Patent No. 6426403
GENERAL INFORMATION:
APPLICANT: NAKATA, MOTOMI
APPLICANT: NAKANO, HIROYASU
APPLICANT: YAGITA, HIDEO
APPLICANT: OKUMURA, KO
TITLE OF INVENTION: TRAP FAMILY MOLECULES, POLYNUCLEOTIDES ENCODING THEM,
AND ANTIBODIES AGAINST THEM
FILE REFERENCE: 007898-0255515
CURRENT FILING DATE: 1998-08-18
PRIOR FILING DATE: 1997-02-24
PRIOR APPLICATION NUMBER: JP 34674/1996
PRIOR FILING DATE: 1996-02-22
NUMBER OF SEQ ID NOS: 16
SOFTWARE: Patent In Ver. 2.1
SEQ ID NO 3
LENGTH: 557
TYPE: PRT
ORGANISM: Homo sapiens
US-09-138-277C-3

Query Match 1.3%; Score 7; DB 4; Length 557;
Best Local Similarity 100.0%; Pred. No. 4.2e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 406 IKKLEKE 412
Db 285 IKKLEKE 291

RESULT 52
US-08-285-440-6
Sequence 6, Application US/08285440
Patent No. 5532337
GENERAL INFORMATION:

APPLICANT: Ken'ichiro HAYASHI et al.
TITLE OF INVENTION: POLYPEPTIDE
NUMBER OF SEQUENCES: 32
CORRESPONDENCE ADDRESS:
ADDRESSEE: Wenderoth, Lind & Ponack
STREET: 805 Fifteenth Street, N.W., #700
CITY: Washington
STATE: D.C.
COUNTRY: U.S.A.
ZIP: 20005
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette, 5.25 inch, 500 kb
COMPUTER: IBM Compatible
OPERATING SYSTEM: MS-DOS
SOFTWARE: DisplayWrite
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/285,440
FILING DATE:
CLASSIFICATION: 536
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 07/859,947
FILING DATE: March 27, 1992
ATTORNEY/AGENT INFORMATION:
NAME: Warren M. Cheek, Jr.
REGISTRATION NUMBER: 33,367
REFERENCE/DOCKET NUMBER:
TELECOMMUNICATION INFORMATION:
TELEPHONE: 202-371-8850
TELEFAX:
TELEX:
INFORMATION FOR SEQ ID NO: 6:
SEQUENCE CHARACTERISTICS:
LENGTH: 558 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: peptide
HYPOTHETICAL:
ANTI-SENSE:
FRAGMENT TYPE:
ORIGINAL SOURCE:
ORGANISM:
STRAIN:
INDIVIDUAL ISOLATE:
DEVELOPMENTAL STAGE:
HAPLOTYPE:
TISSUE TYPE:
CELL TYPE:
CELL LINE:
ORGANELLE:
IMMEDIATE SOURCE:
LIBRARY:
CLONE:
POSITION IN GENOME:
CHROMOSOME/SEGMENT:
MAP POSITION:
UNITS:
FEATURE:
NAME/KEY:
LOCATION:
IDENTIFICATION METHOD:
OTHER INFORMATION:
PUBLICATION INFORMATION:
AUTHORS:
TITLE:
JOURNAL:
VOLUME:
ISSUE:
PAGES:
DATE:
DOCUMENT NUMBER:
FILING DATE:
PUBLICATION DATE:

RELEVANT RESIDUES IN SEQ ID NO:
US-08-285-440-6
Query Match 1.3%; Score 7; DB 1; Length 558;
Best Local Similarity 100.0%; Pred. No. 4.2e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0;
Gaps 0;
QY 248 REEEKR 254
Db 361 REEEKR 367
RESULT 53
US-08-630-349-6
; Sequence 6, Application US/08630349
; Patent No. 5739008
; GENERAL INFORMATION:
; APPLICANT: Ken'ichiro HAYASHI et al.
; TITLE OF INVENTION: POLYPEPTIDE
; NUMBER OF SEQUENCES: 32
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Wenderoth, Lind & Ponack
; STREET: 805 Fifteenth Street, N.W., #700
; CITY: Washington
; STATE: D.C.
; COUNTRY: U.S.A.
; ZIP: 20005
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.5 inch, 144 mb
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: MS-DOS
; SOFTWARE: Wordperfect 5.1
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/630,349
; FILING DATE: April 10, 1996
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/285,440
; FILING DATE: August 4, 1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 07/858,947
; FILING DATE: March 27, 1992
; ATTORNEY/AGENT INFORMATION:
; NAME: Warren M. Cheek, Jr.
; REGISTRATION NUMBER: 33,367
; INFORMATION FOR SEQ ID NO: 6:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 558 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
US-08-630-349-6

Query Match 1.3%; Score 7; DB 1; Length 558;
Best Local Similarity 100.0%; Pred. No. 4.2e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0;
Gaps 0;
QY 248 REEEKR 254
Db 361 REEEKR 367
RESULT 54
US-09-499-522-18
; Sequence 18, Application US/09499522
; Patent No. 6479238
; GENERAL INFORMATION:
; APPLICANT: Blumenfeld, Marta
; APPLICANT: Bougueleret, Lydie
; APPLICANT: Bihain, Bernard
; TITLE OF INVENTION: POLYMORPHIC MARKERS OF THE LSR GENE
; FILE REFERENCE: GENSET.053AUS

; CURRENT APPLICATION NUMBER: US/09/499,522
; CURRENT FILING DATE: 2000-02-10
; EARLIER APPLICATION NUMBER: US 60/119,592
; EARLIER FILING DATE: 1999-02-10
; EARLIER APPLICATION NUMBER: US 60/144,784
; EARLIER FILING DATE: 1999-07-20
; NUMBER OF SEQ ID NOS: 20
; SOFTWARE: Patent.pm
; SEQ ID NO 18
; LENGTH: 581
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: VARIANT
; LOCATION: 295
; OTHER INFORMATION: 9-7-325 : polymorphic amino acid Ser or Asn
; FEATURE:
; NAME/KEY: VARIANT
; LOCATION: 352
; OTHER INFORMATION: 9-9-246 : polymorphic amino acid Pro or Arg
; FEATURE:
; NAME/KEY: VARIANT
; LOCATION: 295
; OTHER INFORMATION: 9-7-325 : polymorphic amino acid deletion of Arg

US-09-499-522-18
Query Match 1.3%; Score 7; DB 4; Length 581;
Best Local Similarity 100.0%; Pred. No. 4.3e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 473 SAGGQGS 479
Db 297 SAGGQGS 303
|||||
|||

RESULT 55
US-09-269-939A-12
; Sequence 12, Application US/09269939A
; Patent No. 6635431
; GENERAL INFORMATION:
; APPLICANT: Bihaun, Bernard
; APPLICANT: Bougueleret, Lydie
; APPLICANT: Yen-Potin, Frances
; TITLE OF INVENTION: LSR Receptor, Its Activity, Its Cloning and Its Application To
; TITLE OF INVENTION: The Diagnosis Prevention And/or Treatment of Obesity and
; TITLE OF INVENTION: Related Risks or Complications
; FILE REFERENCE: GENSET.035APC
; CURRENT APPLICATION NUMBER: US/09/269,939A
; CURRENT FILING DATE: 1999-05-28
; PRIOR APPLICATION NUMBER: FR 97/10088
; PRIOR FILING DATE: 1997-08-06
; PRIOR APPLICATION NUMBER: FR 98/05032
; PRIOR FILING DATE: 1998-04-22
; NUMBER OF SEQ ID NOS: 41
; SOFTWARE: Patent.pm
; SEQ ID NO 12
; LENGTH: 581
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-269-939A-12

Query Match 1.3%; Score 7; DB 4; Length 581;
Best Local Similarity 100.0%; Pred. No. 4.3e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 473 SAGGQGS 479
Db 297 SAGGQGS 303
|||||
|||

RESULT 56
US-09-252-991A-27626
; Sequence 27626, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 107196.1136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 27626
; LENGTH: 582
; TYPE: PRT
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-27626

Query Match 1.3%; Score 7; DB 4; Length 630;
Best Local Similarity 100.0%; Pred. No. 4.6e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 473 SAGGQGS 479
Db 346 SAGGQGS 352
|||||
|||

Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 107196.1136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 27626
; LENGTH: 582
; TYPE: PRT
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-27626

Query Match 1.3%; Score 7; DB 4; Length 582;
Best Local Similarity 100.0%; Pred. No. 4.3e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 155 EITLLMQ 161
Db 355 EITLLMQ 361
|||||
|||

RESULT 57
US-09-499-522-16
; Sequence 16, Application US/09499522
; Patent No. 6479238
; GENERAL INFORMATION:
; APPLICANT: Blumenfeld, Marta
; APPLICANT: Bougueleret, Lydie
; APPLICANT: Bihaun, Bernard
; TITLE OF INVENTION: POLYMORPHIC MARKERS OF THE LSR GENE
; FILE REFERENCE: GENSET.053AUS
; CURRENT APPLICATION NUMBER: US/09/499,522
; CURRENT FILING DATE: 2000-02-10
; EARLIER APPLICATION NUMBER: US 60/119,592
; EARLIER FILING DATE: 1999-02-10
; EARLIER APPLICATION NUMBER: US 60/144,784
; EARLIER FILING DATE: 1999-07-20
; NUMBER OF SEQ ID NOS: 20
; SOFTWARE: Patent.pm
; SEQ ID NO 16
; LENGTH: 630
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: VARIANT
; LOCATION: 344
; OTHER INFORMATION: 9-7-325 : polymorphic amino acid Ser or Asn
; FEATURE:
; NAME/KEY: VARIANT
; LOCATION: 401
; OTHER INFORMATION: 9-9-246 : polymorphic amino acid Pro or Arg
; FEATURE:
; NAME/KEY: VARIANT
; LOCATION: 500
; OTHER INFORMATION: LSRX9f13-BM : polymorphic amino acid deletion of Arg
US-09-499-522-16

Query Match 1.3%; Score 7; DB 4; Length 630;
Best Local Similarity 100.0%; Pred. No. 4.6e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 473 SAGGQGS 479
Db 346 SAGGQGS 352
|||||
|||

```

RESULT 58
US-09-269-939A-10
; Sequence 10, Application US/09269939A
; Patent No. 6635431
; GENERAL INFORMATION:
; APPLICANT: Bihaïn, Bernard
; APPLICANT: Bougueleret, Lydie
; TITLE OF INVENTION: LSR Receptor, Its Activity, Its Cloning and Its Application To
; TITLE OF INVENTION: The Diagnosis Prevention And/or Treatment of Obesity and
; TITLE OF INVENTION: Related Risks or Complications
; FILE REFERENCE: GENSET.035APC
; CURRENT APPLICATION NUMBER: US/09/269,939A
; CURRENT FILING DATE: 1999-05-28
; PRIOR APPLICATION NUMBER: FR 97/10088
; PRIOR FILING DATE: 1997-08-06
; PRIOR APPLICATION NUMBER: FR 98/05032
; PRIOR FILING DATE: 1998-04-22
; NUMBER OF SEQ ID NOS: 41
; SOFTWARE: Patent.pm
; SEQ ID NO 10
; LENGTH: 630
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-269-939A-10

Query Match 1.3%; Score 7; DB 4; Length 630;
Best Local Similarity 100.0%; Pred. No. 4.6e+02; Indels 0; Gaps 0;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 473 SAGGQGS 479
DB 346 SAGGQGS 352

RESULT 59
US-09-499-522-14
; Sequence 14, Application US/09499522
; Patent No. 6479238
; GENERAL INFORMATION:
; APPLICANT: Blumenfeld, Marta
; APPLICANT: Bougueleret, Lydie
; APPLICANT: Bihaïn, Bernard
; TITLE OF INVENTION: POLYMORPHIC MARKERS OF THE LSR GENE
; FILE REFERENCE: GENSET.053AUS
; CURRENT APPLICATION NUMBER: US/09/499,522
; CURRENT FILING DATE: 2000-02-10
; EARLIER APPLICATION NUMBER: US 60/119,592
; EARLIER FILING DATE: 1999-02-10
; EARLIER APPLICATION NUMBER: US 60/144,784
; EARLIER FILING DATE: 1999-07-20
; NUMBER OF SEQ ID NOS: 20
; SOFTWARE: Patent.pm
; SEQ ID NO 14
; LENGTH: 649
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: VARIANT
; LOCATION: 363
; OTHER INFORMATION: 9-7-325 : polymorphic amino acid Ser or Asn
; FEATURE:
; NAME/KEY: VARIANT
; LOCATION: 420
; OTHER INFORMATION: 9-9-246 : polymorphic amino acid Pro or Arg
; FEATURE:
; NAME/KEY: VARIANT
; LOCATION: 519
; OTHER INFORMATION: LSRX9f13-BM : polymorphic amino acid deletion of Arg
US-09-499-522-14

Query Match 1.3%; Score 7; DB 4; Length 649;
Best Local Similarity 100.0%; Pred. No. 4.7e+02; Indels 0; Gaps 0;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 473 SAGGQGS 479
DB 346 SAGGQGS 352
```

```

Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 473 SAGGQGS 479
DB 365 SAGGQGS 371

RESULT 60
US-09-269-939A-8
; Sequence 8, Application US/09269939A
; Patent No. 6635431
; GENERAL INFORMATION:
; APPLICANT: Bihaïn, Bernard
; APPLICANT: Bougueleret, Lydie
; APPLICANT: Yen-Potin, Frances
; TITLE OF INVENTION: LSR Receptor, Its Activity, Its Cloning and Its Application To
; TITLE OF INVENTION: The Diagnosis Prevention And/or Treatment of Obesity and
; TITLE OF INVENTION: Related Risks or Complications
; FILE REFERENCE: GENSET.035APC
; CURRENT APPLICATION NUMBER: US/09/269,939A
; CURRENT FILING DATE: 1999-05-28
; PRIOR APPLICATION NUMBER: FR 97/10088
; PRIOR FILING DATE: 1997-08-06
; PRIOR APPLICATION NUMBER: FR 98/05032
; PRIOR FILING DATE: 1998-04-22
; NUMBER OF SEQ ID NOS: 41
; SOFTWARE: Patent.pm
; SEQ ID NO 8
; LENGTH: 649
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: Misc_Feature
; LOCATION: 386
; OTHER INFORMATION: Potential deletion of a Glu
; NAME/KEY: Misc_Feature
; LOCATION: 518
; OTHER INFORMATION: Potential insertion of a Arg
US-09-269-939A-8

Query Match 1.3%; Score 7; DB 4; Length 649;
Best Local Similarity 100.0%; Pred. No. 4.7e+02; Indels 0; Gaps 0;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 473 SAGGQGS 479
DB 365 SAGGQGS 371

RESULT 61
US-09-543-681A-5392
; Sequence 5392, Application US/09543681A
; Patent No. 6605709
; GENERAL INFORMATION:
; APPLICANT: GARY BRETON
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PROTEUS MIRABILIS
; FILE REFERENCE: 2709.1002-001
; CURRENT APPLICATION NUMBER: US/09/543,681A
; CURRENT FILING DATE: 2000-04-05
; PRIOR APPLICATION NUMBER: US 60/128,706
; PRIOR FILING DATE: 1999-04-09
; NUMBER OF SEQ ID NOS: 8344
; SEQ ID NO 5392
; LENGTH: 712
; TYPE: PRT
; ORGANISM: Proteus mirabilis
US-09-543-681A-5392

Query Match 1.3%; Score 7; DB 4; Length 712;
Best Local Similarity 100.0%; Pred. No. 5.1e+02; Indels 0; Gaps 0;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

QY 46 TAQSGAL 52
|||||
Db 373 TAQSGAL 379

RESULT 62
US-09-562-737-47
; Sequence 47, Application US/09562737
; Patent No. 6428967
; GENERAL INFORMATION:
; APPLICANT: Herz, Joachim
; APPLICANT: Gotthardt, Michael
; TITLE OF INVENTION: LDL Receptor Signaling Pathways
; FILE REFERENCE: UTSW0708
; CURRENT APPLICATION NUMBER: US/09/562,737
; CURRENT FILING DATE: 2000-05-01
; NUMBER OF SEQ ID NOS: 132
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 47
; LENGTH: 1024
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: Synthetic
; OTHER INFORMATION: Sequence
US-09-562-737-47

Query Match 1.3%; Score 7; DB 4; Length 1024;
Best Local Similarity 100.0%; Pred. No. 6.8e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 343 KDFLLKE 349
|||||
Db 263 KDFLLKE 269

RESULT 63
US-09-252-991A-18387
; Sequence 18387, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; FILE REFERENCE: 107195.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 18387
; LENGTH: 1075
; TYPE: PRT
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-18387

Query Match 1.3%; Score 7; DB 4; Length 1075;
Best Local Similarity 100.0%; Pred. No. 7.1e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 20 QAAPAVE 26
|||||
Db 999 QAAPAVE 1005

RESULT 64
US-08-131-365B-54
; Sequence 54, Application US/08131365B
; Patent No. 5527690
; GENERAL INFORMATION:
; APPLICANT: Brown, Michael S.

; APPLICANT: Briggs, Michael R.
; APPLICANT: Wang, Xiaodong
; APPLICANT: Goldstein, Joseph L.
; TITLE OF INVENTION: METHODS AND COMPOSITIONS RELATING
; TITLE OF INVENTION: TO STEROL REGULATORY ELEMENT BINDING
; TITLE OF INVENTION: PROTEINS
; NUMBER OF SEQUENCES: 64
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Arnold, White & Durkee
; STREET: P.O. Box 4433
; CITY: Houston
; STATE: Texas
; COUNTRY: U.S.A.
; ZIP: 77210
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA: US/08/131,365B
; APPLICATION NUMBER: US/08/131,365B
; FILING DATE: 01-OCT-1993
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Parker, David L.
; REGISTRATION NUMBER: 32,165
; REFERENCE/DOCKET NUMBER: UTSD:372/PAR
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (512) 418-3000
; TELEFAX: (512) 474-7577
; INFORMATION FOR SEQ ID NO: 54:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1141
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-131-365B-54

Query Match 1.3%; Score 7; DB 1; Length 1141;
Best Local Similarity 100.0%; Pred. No. 7.5e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 281 KLRQENM 287
|||||
Db 386 KLRQENM 392

RESULT 65
US-08-668-123-54
; Sequence 54, Application US/08668123
; Patent No. 5891631
; GENERAL INFORMATION:
; APPLICANT: Brown, Michael S.
; APPLICANT: Briggs, Michael R.
; APPLICANT: Wang, Xiaodong
; APPLICANT: Goldstein, Joseph L.
; TITLE OF INVENTION: METHODS AND COMPOSITIONS RELATING
; TITLE OF INVENTION: TO STEROL REGULATORY ELEMENT BINDING
; TITLE OF INVENTION: PROTEINS
; NUMBER OF SEQUENCES: 64
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Arnold, White & Durkee
; STREET: P.O. Box 4433
; CITY: Houston
; STATE: Texas
; COUNTRY: U.S.A.
; ZIP: 77210
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30

; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/668,123
; FILING DATE: 14-JUN-1996
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/131,365
; FILING DATE: 01-OCT-1993
; ATTORNEY/AGENT INFORMATION:
; NAME: Parker, David L.
; REGISTRATION NUMBER: 32,165
; REFERENCE/DOCKET NUMBER: UTSD:372/PAR
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (512) 418-3000
; TELEFAX: (512) 474-7577
; INFORMATION FOR SEQ ID NO: 54:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1141
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-668-123-54

Query Match 1.3%; Score 7; DB 2; Length 1141;
Best Local Similarity 100.0%; Pred. No. 7.5e+02; Mismatches 0; Indels 0; Gaps 0;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 281 KLRQNM 287
| | | | |
DB 386 KLRQNM 392

RESULT 66

US-09-976-594-736
; Sequence 736, Application US/09976594
; Patent No. 6673549
; GENERAL INFORMATION:
; APPLICANT: Furness, Michael
; APPLICANT: Buchbinder, Jenny
; TITLE OF INVENTION: GENES EXPRESSED IN C3A LIVER CELL CULTURES TREATED WITH STEROIDS
; FILE REFERENCE: PA-0041 US
; CURRENT APPLICATION NUMBER: US/09/976,594
; CURRENT FILING DATE: 2001-10-12
; PRIOR APPLICATION NUMBER: 60/240,409
; PRIOR FILING DATE: 2000-10-12
; NUMBER OF SEQ ID NOS: 1143
; SOFTWARE: PERL Program
; SEQ ID NO 736
; LENGTH: 1530
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc feature
; OTHER INFORMATION: Incyte ID No. 6673549 3151579CD1
; US-09-976-594-736

Query Match 1.3%; Score 7; DB 4; Length 1530;
Best Local Similarity 100.0%; Pred. No. 9.5e+02; Mismatches 0; Indels 0; Gaps 0;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 439 KELEGLO 445
| | | | |
DB 1316 KELEGLO 1322

RESULT 67

US-08-287-959-1
; Sequence 1, Application US/08287959
; Patent No. 5639651
; GENERAL INFORMATION:
; APPLICANT: Weissbach, Lawrence
; APPLICANT: Bernards, Andre
; APPLICANT: Settleman, Jeffrey

Query Match 1.3%; Score 7; DB 4; Length 3340;
Best Local Similarity 100.0%; Pred. No. 1.8e+03; Mismatches 0; Indels 0; Gaps 0;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 365 THLKQOL 371
| | | | |
DB 2522 THLKQOL 2528

; TITLE OF INVENTION: GAP-RELATED GENE
; NUMBER OF SEQUENCES: 26
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Fish & Richardson
; STREET: 225 Franklin Street
; CITY: Boston
; STATE: MA
; COUNTRY: U.S.A.
; ZIP: 02110
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/287,959
; FILING DATE: August 9, 1994
; CLASSIFICATION: 514
; ATTORNEY/AGENT INFORMATION:
; NAME: Clark, Paul C.
; REGISTRATION NUMBER: 30,162
; REFERENCE/DOCKET NUMBER: 00786/181001
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617) 542-5070
; TELEFAX: (617) 542-8906
; TELEX: 200154
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1657 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-287-959-1

Query Match 1.3%; Score 7; DB 1; Length 1657;
Best Local Similarity 100.0%; Pred. No. 1e+03; Mismatches 0; Indels 0; Gaps 0;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 471 DLSAGGQ 477
| | | | |
DB 1210 DLSAGGQ 1216

RESULT 68

US-09-252-991A-23568
; Sequence 23568, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 23568
; LENGTH: 3340
; TYPE: PRT
; ORGANISM: Pseudomonas aeruginosa
; US-09-252-991A-23568

Query Match 1.3%; Score 7; DB 4; Length 3340;
Best Local Similarity 100.0%; Pred. No. 1.8e+03; Mismatches 0; Indels 0; Gaps 0;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 365 THLKQOL 371
| | | | |
DB 2522 THLKQOL 2528

RESULT 69
US-08-310-912A-138
; Sequence 138, Application US/08310912A
; Patent No. 5981730
; GENERAL INFORMATION:
; APPLICANT: Ausubel, Frederick M.
; APPLICANT: Staskawicz, Brian J.
; APPLICANT: Brent, Andrew F.
; APPLICANT: Dahlbeck, Douglas
; APPLICANT: Katagiri, Fumiaki
; APPLICANT: Kunkel, Barbara N.
; APPLICANT: Mindrinos, Michael N.
; APPLICANT: Yu, Guo-Liang
; TITLE OF INVENTION: RPS2 GENE FAMILY, PRIMERS, PROBES, AND DETECTION
; NUMBER OF SEQUENCES: 208
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Fish & Richardson P.C.
; STREET: 225 Franklin Street
; CITY: Boston
; STATE: MA
; COUNTRY: USA
; ZIP: 02110-2904
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30B
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/310.912A
; FILING DATE: September 22, 1994
; CLASSIFICATION: 536
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/227,360
; FILING DATE: April 13, 1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Lech, Karen F.
; REGISTRATION NUMBER: 35,238
; REFERENCE/DOCKET NUMBER: 00786/254001
; TELEPHONE: (617) 542-5070
; TELEFAX: (617) 542-8906
; TELEX: 100254
; INFORMATION FOR SEQ ID NO: 138:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 6 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-310-912A-138

Query Match 1.1%; Score 6; DB 2; Length 6;
Best Local Similarity 100.0%; Pred. No. 3e+05;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 223 SKLESL 228
Db 1 SKLESL 6

RESULT 71
PCT-US95-04589-138
; Sequence 138, Application PC/TUS9504589
; GENERAL INFORMATION:
; APPLICANT: Ausubel, Frederick M.
; APPLICANT: Staskawicz, Brian J.
; APPLICANT: Brent, Andrew F.
; APPLICANT: Dahlbeck, Douglas
; APPLICANT: Katagiri, Fumiaki
; APPLICANT: Kunkel, Barbara N.
; APPLICANT: Mindrinos, Michael N.
; APPLICANT: Yu, Guo-Liang
; TITLE OF INVENTION: RPS2 GENE AND USBS THEREOF
; NUMBER OF SEQUENCES: 201
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Fish & Richardson
; STREET: 225 Franklin Street Suite 3100
; CITY: Boston
; STATE: MA
; COUNTRY: USA
; ZIP: 02110-2904
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30B
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US95/04589
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/227,360
; FILING DATE: 13-APR-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Clark, Paul T.
; REGISTRATION NUMBER: 30,162
; REFERENCE/DOCKET NUMBER: 00786/230001
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617) 542-5070
; TELEFAX: (617) 542-8906
; TELEX: 100254
; INFORMATION FOR SEQ ID NO: 138:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 6 amino acids
; TYPE: amino acid
; TOPOLOGY: linear

```

; MOLECULE TYPE: protein
PCT-US95-04589-138

Query Match      1.1%; Score 6; DB 5; Length 6;
Best Local Similarity 100.0%; Pred. No. 3e+05;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 223 SKLESL 228
DB 1 SKLESL 6

RESULT 72
US-09-989-789-1988
; Sequence 1988, Application US/09989789
; Patent No. 8588746
; GENERAL INFORMATION:
; APPLICANT: LIU, Qiang
; TITLE OF INVENTION: POSITION DEPENDENT RECOGNITION OF GNN NUCLEOTIDE
; TITLE OF INVENTION: TRIPLETS BY ZINC FINGERS
; FILE REFERENCE: 8325-0011.20 / S11-US2
; CURRENT APPLICATION NUMBER: US/09/989,789
; CURRENT FILING DATE: 2002-03-25
; NUMBER OF SEQ ID NOS: 4085
; SOFTWARE: Patentin Ver. 2.0
; SEQ ID NO 1988
; LENGTH: 7
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: example ZFP
US-09-989-789-1988

Query Match      1.1%; Score 6; DB 4; Length 7;
Best Local Similarity 100.0%; Pred. No. 3e+05;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 48 QSGALR 53
DB 1 QSGALR 6

RESULT 73
US-08-339-152A-1
; Sequence 1, Application US/08339152A
; Patent No. 5643726
; GENERAL INFORMATION:
; APPLICANT: Tanzi, Rudolph E.
; APPLICANT: Kovacs, Dora M.
; TITLE OF INVENTION: Methods For Modulating Transcription
; TITLE OF INVENTION: From The Amyloid -Protein Precursor (APP) Promoter
; NUMBER OF SEQUENCES: 33
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: STERN, KESSLER, GOLDSTEIN & FOX P.L.L.C.
; STREET: 1100 New York Ave., NW, Suite 600
; CITY: Washington
; STATE: DC
; COUNTRY: USA
; ZIP: 20005
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/339,152A
; FILING DATE: 10-NOV-1994
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Steffe, Eric K.
; REGISTRATION NUMBER: 36,688
; REFERENCE/DOCKET NUMBER: 0609.4120000
; TELECOMMUNICATION INFORMATION:

Query Match      1.1%; Score 6; DB 2; Length 9;
Best Local Similarity 100.0%; Pred. No. 3e+05;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 230 RELORH 235
DB 4 RELORH 9

RESULT 74
US-08-007-999B-1
; Sequence 1, Application US/08007999B
; Patent No. 5851787
; GENERAL INFORMATION:
; APPLICANT: Wasco, Wilma
; APPLICANT: Bupp, Keith
; APPLICANT: Magendantz, Margaret
; APPLICANT: Tanzi, Rudolph
; APPLICANT: Solomon, Frank
; TITLE OF INVENTION: AMYLOID PRECURSOR-LIKE PROTEIN AND USES
; TITLE OF INVENTION: THEREOF
; NUMBER OF SEQUENCES: 19
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Sterne, Kessler, Goldstein & Fox, P.L.L.C.
; STREET: 1100 New York Ave., NW
; CITY: Washington
; STATE: DC
; COUNTRY: USA
; ZIP: 20005
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/007,999B
; FILING DATE: 21-JAN-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/872,642
; FILING DATE: 20-APR-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/930,022
; FILING DATE: 17-AUG-1992
; ATTORNEY/AGENT INFORMATION:
; NAME: Townsend, G. Kevin
; REGISTRATION NUMBER: 34,033
; REFERENCE/DOCKET NUMBER: 0609.3520002/JAG/GKT
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202)371-2571
; TELEFAX: (202)371-2540
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 9 amino acids
; TYPE: amino acid
; STRANDEDNESS: both
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
US-08-007-999B-1
```

QY 230 RELQKH 235
|||||
Db 4 RELQKH 9

RESULT 75

US-08-689-276A-1
; Sequence 1, Application US/08689276A
; Patent No. 5891991
; GENERAL INFORMATION:
; APPLICANT: Masco, Wilma
; APPLICANT: Bupp, Keith
; APPLICANT: Magendantz, Margaret
; APPLICANT: Tanzi, Rudolph
; APPLICANT: Solomon, Frank
; TITLE OF INVENTION: AMYLOID PRECURSOR-LIKE PROTEIN AND USES
; TITLE OF INVENTION: THEREOF
; NUMBER OF SEQUENCES: 19
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Sterne, Kessler, Goldstein & Fox P.L.L.C.
; STREET: 1100 New York Ave., NW
; CITY: Washington
; STATE: DC
; COUNTRY: USA
; ZIP: 20005
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/689,276A
; FILING DATE: 06-AUG-1996
; PRIOR APPLICATION NUMBER:
; FILING DATE: 21-JAN-1993
; APPLICATION NUMBER: US 08/007,999
; FILING DATE: 21-JAN-1993
; PRIOR APPLICATION NUMBER: US 07/872,642
; FILING DATE: 20-APR-1992
; APPLICATION NUMBER: US 07/930,022
; FILING DATE: 17-AUG-1992
; ATTORNEY/AGENT INFORMATION:
; NAME: JORGE A. GOLDSTEIN
; REGISTRATION NUMBER: 29,021
; REFERENCES/DOCKET NUMBER: 0609.3520003
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202)371-2571
; TELEFAX: (202)371-2540
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 9 amino acids
; TYPE: amino acid
; STRANDEDNESS: both
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
US-08-689-276A-1

Query Match 1.1%; Score 6; DB 2; Length 9;
Best Local Similarity 100.0%; Pred. No. 3e+05; 0; Indels 0; Gaps 0;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 230 RELQKH 235
|||||
Db 4 RELQKH 9

RESULT 76

US-08-159-339A-830
; Sequence 830, Application US/08159339A
; Patent No. 6037135
; GENERAL INFORMATION:
; APPLICANT: Kubo, Ralph T.

; APPLICANT: Grey, Howard M.
; APPLICANT: Sette, Alessandro
; APPLICANT: Celis, Esben
; TITLE OF INVENTION: HLA Binding peptides and Their
; TITLE OF INVENTION: Uses
; NUMBER OF SEQUENCES: 1254
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Townsend and Townsend and Crew LLP
; STREET: Two Embarcadero Center, Eighth Floor
; CITY: San Francisco
; STATE: CA
; COUNTRY: USA
; ZIP: 94111-3834
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/159,339A
; FILING DATE: 29-NOV-1993
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/925,666
; FILING DATE: 07-AUG-1992
; APPLICATION NUMBER: US 08/027,746
; FILING DATE: 05-MAR-1993
; APPLICATION NUMBER: US 08/103,396
; FILING DATE: 06-AUG-1993
; ATTORNEY/AGENT INFORMATION:
; NAME: Weber, Ellen Lauver
; REGISTRATION NUMBER: 32,762
; REFERENCES/DOCKET NUMBER: 018623-005030US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 576-0200
; TELEFAX: (415) 576-0300
; TELEX:
; INFORMATION FOR SEQ ID NO: 830:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 10 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
US-08-159-339A-830

Query Match 1.1%; Score 6; DB 3; Length 10;
Best Local Similarity 100.0%; Pred. No. 1.3e+02; 0; Indels 0; Gaps 0;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 10 GPEGCAQ 15
|||||
Db 2 GPEGCAQ 7

RESULT 77

US-08-836-075A-109
; Sequence 109, Application US/08836075A
; Patent No. 6180768
; GENERAL INFORMATION:
; APPLICANT: MAERTENS, GERT
; APPLICANT: STUYVER, LIEVEN
; TITLE OF INVENTION: NEW SEQUENCES OF HEPATITIS C VIRUS GENOTYPES
; TITLE OF INVENTION: AND THEIR USE AS PROPHYLACTIC, THERAPEUTIC AND DIAGNOSTIC
; TITLE OF INVENTION: AGENTS
; NUMBER OF SEQUENCES: 207
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: ARNOLD, WHITE & DURKEE
; STREET: P.O. BOX 4433
; CITY: HOUSTON
; STATE: TEXAS
; COUNTRY: USA
; ZIP: 77210-4433


```
;
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Microsoft Word 6.0 / ASCII text output
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/836,075A
; FILING DATE: 21 Apr 1997
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PCT/EP95/04155
; FILING DATE: 23 Oct 1995
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: EP 94870166.9
; FILING DATE: 21 Oct 1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: EP 95870076.7
; FILING DATE: 28 Jun 1995
; ATTORNEY/AGENT INFORMATION:
; NAME: KAMMERER, PATRICIA A.
; REGISTRATION NUMBER: 29,775
; REFERENCE/DOCKET NUMBER: INNS:004
; INFORMATION FOR SEQ ID NO: 109:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 11 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; US-08-836-075A-109

Query Match 1.1%; Score 6; DB 3; Length 11;
Best Local Similarity 100.0%; Pred. No. 1.4e+02;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 486 ERRPEG 491
DB 1 ERRPEG 6

RESULT 78
US-08-747-599A-16
; Sequence 16, Application US/08747599A
; Patent No. 6214795
; GENERAL INFORMATION:
; APPLICANT: Benjamin, Howard et al.
; TITLE OF INVENTION: Peptide Compounds Useful for Modulating
; TITLE OF INVENTION: FGF Receptor Activity
; NUMBER OF SEQUENCES: 68
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: LAHIVE & COCKFIELD
; STREET: 28 State Street
; CITY: Boston
; STATE: Massachusetts
; COUNTRY: USA
; ZIP: 02109-1875
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/747,599A
; FILING DATE: 12-NOV-1996
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Kara, Catherine J.
; REGISTRATION NUMBER: 41,106
; REFERENCE/DOCKET NUMBER: PPI-029
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617)227-7400
; TELEFAX: (617)227-5941
; INFORMATION FOR SEQ ID NO: 16:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 12 amino acids

;
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; FRAGMENT TYPE: internal
; US-08-747-599A-16

Query Match 1.1%; Score 6; DB 3; Length 12;
Best Local Similarity 100.0%; Pred. No. 1.5e+02;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 371 LALYTE 376
DB 4 LALYTE 9

RESULT 79
US-08-979-608A-29
; Sequence 29, Application US/08979608A
; Patent No. 6355451
; GENERAL INFORMATION:
; APPLICANT: Lees, Ann M.
; Lees, Robert S.
; Law, Simon W.
; Arjona, Anibal A.
; TITLE OF INVENTION: NOVEL LOW DENSITY LIPOPROTEIN
; BINDING PROTEINS AND THEIR USES IN DIAGNOSING AND
; TREATING ATHEROSCLEROSIS
; NUMBER OF SEQUENCES: 42
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Fish & Richardson P.C.
; STREET: 225 Franklin Street
; CITY: Boston
; STATE: MA
; COUNTRY: USA
; ZIP: 02110-2804
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSEQ for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/979,608A
; FILING DATE: 26-No. 6355451-1997
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 60/048,547
; FILING DATE: 03-JUN-1997
; APPLICATION NUMBER: US 60/031,930
; FILING DATE: 27-NOV-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: Myers, Louis
; REGISTRATION NUMBER: 35,965
; REFERENCE/DOCKET NUMBER: 10797-002001 (formerly 3983/59818)
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 617/542-5070
; TELEFAX: 617/542-8906
; INFORMATION FOR SEQ ID NO: 29:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 15 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; SEQUENCE DESCRIPTION: SEQ ID NO: 29:
; US-08-979-608A-29

Query Match 1.1%; Score 6; DB 4; Length 15;
Best Local Similarity 100.0%; Pred. No. 1.8e+02;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 90 EPEDAE 95
DB 10 EPEDAE 15
```

RESULT 80

US-09-517-849-29
; Sequence 29, Application US/09517849
; Patent No. 6605588
; GENERAL INFORMATION:
; APPLICANT: Lees, Ann M.
; Lees, Robert S.
; Law, Simon W.
; Arjona, Anibal A.
; TITLE OF INVENTION: NOVEL LOW DENSITY LIPOPROTEIN
; BINDING PROTEINS AND THEIR USES IN DIAGNOSING AND
; TREATING ATHEROSCLEROSIS
; NUMBER OF SEQUENCES: 42
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Fish & Richardson P.C.
; STREET: 225 Franklin Street
; CITY: Boston
; STATE: MA
; COUNTRY: USA
; ZIP: 02110-2804
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSEQ for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/517,849
; FILING DATE: 02-Mar-2000
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/979,608
; FILING DATE: 26-NOV-1997
; ATTORNEY/AGENT INFORMATION:
; NAME: Myers, Louis
; REGISTRATION NUMBER: 35,965
; REFERENCE/DOCKET NUMBER: 10797-003001
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 617/542-5070
; TELEFAX: 617/542-8906
; INFORMATION FOR SEQ ID NO: 29:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 15 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; SEQUENCE DESCRIPTION: SEQ ID NO: 29:

US-09-517-849-29

Query Match 1.1%; Score 6; DB 4; Length 15;
Best Local Similarity 100.0%; Pred. No. 1.8e+02;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 90 EPEDAE 95

DB 10 EPEDAE 15

RESULT 81

US-09-616-289-29
; Sequence 29, Application US/09616289
; Patent No. 6632923
; GENERAL INFORMATION:
; APPLICANT: Lees, Ann M.
; Lees, Robert S.
; Law, Simon W.
; APPLICANT: Arjona, Anibal A.
; TITLE OF INVENTION: NOVEL LOW DENSITY LIPOPROTEIN BINDING
; PROTEINS AND THEIR USE IN DIAGNOSING AND TREATING
; ATHEROSCLEROSIS
; TITLE OF INVENTION: ATHEROSCLEROSIS
; FILE REFERENCE: 10797-004001
; CURRENT APPLICATION NUMBER: US/09/616,289
; CURRENT FILING DATE: 2000-07-14
; PRIOR APPLICATION NUMBER: US 09/517,849
; PRIOR FILING DATE: 2000-03-02

; PRIOR APPLICATION NUMBER: US 08/979,608
; PRIOR FILING DATE: 1997-11-26
; PRIOR APPLICATION NUMBER: US 60/031,930
; PRIOR FILING DATE: 1996-11-27
; PRIOR APPLICATION NUMBER: US 60/048,547
; PRIOR FILING DATE: 1997-06-03
; NUMBER OF SEQ ID NOS: 53
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 29
; LENGTH: 15
; TYPE: PRT
; ORGANISM: Oryctolagus cuniculus
; US-09-616-289-29

Query Match 1.1%; Score 6; DB 4; Length 15;
Best Local Similarity 100.0%; Pred. No. 1.8e+02;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 90 EPEDAE 95

DB 10 EPEDAE 15

RESULT 82

US-08-408-120-15
; Sequence 15, Application US/08408120
; Patent No. 5804200
; GENERAL INFORMATION:
; APPLICANT: GRIEVE, ROBERT B.
; APPLICANT: FRANK, GLENN R.
; TITLE OF INVENTION: REAGENTS AND METHODS FOR IDENTIFICATION
; OF VACCINES
; NUMBER OF SEQUENCES: 30
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: MORRISON & FOERSTER
; STREET: 755 Page Mill Road
; CITY: Palo Alto
; STATE: California
; COUNTRY: USA
; ZIP: 94304-1018
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/408,120
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/003,257
; FILING DATE: 12-JAN-1993
; ATTORNEY/AGENT INFORMATION:
; NAME: MURASHIGE, KATE H.
; REGISTRATION NUMBER: 29,959
; REFERENCE/DOCKET NUMBER: 27010-20006.20
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 813-5600
; TELEFAX: (415) 494-0792
; TELEX: 706141
; INFORMATION FOR SEQ ID NO: 15:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 16 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; US-08-408-120-15

Query Match 1.1%; Score 6; DB 1; Length 16;
Best Local Similarity 100.0%; Pred. No. 1.9e+02;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 83 GAGEEP 88

Page 25

Db 13 EEBKR 18

US-09-834-759-544

Sequence 544, Application US/09834759

Patent No. 6680197

GENERAL INFORMATION:

APPLICANT: Jiang, Yugu

APPLICANT: Dillon, Davin C.

APPLICANT: Mitcham, Jennifer L.

APPLICANT: Xu, Jiangchun

APPLICANT: Harlocker, Susan L.

APPLICANT: Hepler, William T.

APPLICANT: Henderson, Robert A.

TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY AND

FILE REFERENCE: 210121.470C9

CURRENT FILING DATE: 2001-04-13

NUMBER OF SEQ ID NOS: 547

SOFTWARE: FastSeq for Windows Version 3.0

SEQ ID NO 544

LENGTH: 20

TYPE: PRT

ORGANISM: Homo sapiens

US-09-834-759-544

Query Match 1.1%; Score 6; DB 4; Length 20;

Best Local Similarity 100.0%; Pred. No. 2.2e+02; Indels 0; Gaps 0;

Matches 6; Conservative 0; Mismatches 0;

QY 316 LQQQLV 321

Db 5 LQQQLV 10

US-09-834-759-543

Sequence 543, Application US/09834759

Patent No. 6680197

GENERAL INFORMATION:

APPLICANT: Jiang, Yugu

APPLICANT: Dillon, Davin C.

APPLICANT: Mitcham, Jennifer L.

APPLICANT: Xu, Jiangchun

APPLICANT: Harlocker, Susan L.

APPLICANT: Hepler, William T.

APPLICANT: Henderson, Robert A.

TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY AND

FILE REFERENCE: 210121.470C9

CURRENT FILING DATE: 2001-04-13

NUMBER OF SEQ ID NOS: 547

SOFTWARE: FastSeq for Windows Version 3.0

SEQ ID NO 543

LENGTH: 21

TYPE: PRT

ORGANISM: Homo sapiens

US-09-834-759-543

Query Match 1.1%; Score 6; DB 4; Length 21;

Best Local Similarity 100.0%; Pred. No. 2.3e+02; Indels 0; Gaps 0;

Matches 6; Conservative 0; Mismatches 0;

QY 316 LQQQLV 321

Db 14 LQQQLV 19

US-08-851-843A-113

Sequence 233, Application US/08974549A

Patent No. 6166178

GENERAL INFORMATION:

APPLICANT: Cech, Thomas R.

APPLICANT: Lingner, Joachim

APPLICANT: Nakamura, Toru

APPLICANT: Chapman, Karen B.

APPLICANT: Morin, Gregg B.

APPLICANT: Harley, Calvin

APPLICANT: Andrews, William H.

TITLE OF INVENTION: No. 6033809e1 Telomerase

NUMBER OF SEQUENCES: 225

CORRESPONDENCE ADDRESS:

ADDRESSEE: Townsend and Townsend and Crew LLP

STREET: Two Embarcadero Center, 8th Floor

CITY: San Francisco

STATE: California

COUNTRY: United States of America

ZIP: 94111

COMPUTER READABLE FORM: disk

MEDIUM TYPE: Floppy

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: PatentIn Release #1.0, Version #1.30

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/851,843A

FILING DATE: 06-MAY-1997

CLASSIFICATION:

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 08/846,017

FILING DATE: 25-APR-1997

CLASSIFICATION:

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 08/844,419

FILING DATE: 18-APR-1997

CLASSIFICATION:

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 08/724,643

FILING DATE: 01-OCT-1996

CLASSIFICATION:

ATTORNEY/AGENT INFORMATION:

NAME: Apple, Randolph T.

REGISTRATION NUMBER: 36,429

REFERENCE/DOCKET NUMBER: 015389-002930US

TELECOMMUNICATION INFORMATION:

TELEPHONE: (415) 576-0200

TELEFAX: (415) 576-0300

INFORMATION FOR SEQ ID NO: 113:

SEQUENCE CHARACTERISTICS:

LENGTH: 23 amino acids

TYPE: amino acid

STRANDEDNESS:

TOPOLOGY: linear

MOLECULE TYPE: peptide

US-08-851-843A-113

Query Match 1.1%; Score 6; DB 3; Length 23;

Best Local Similarity 100.0%; Pred. No. 2.5e+02; Indels 0; Gaps 0;

Matches 6; Conservative 0; Mismatches 0;

QY 52 LRDVSE 57

Db 9 LRDVSE 14

US-08-974-549A-233

Sequence 233, Application US/08974549A

Patent No. 6166178

GENERAL INFORMATION:

APPLICANT: Cech, Thomas R.

APPLICANT: Lingner, Joachim

APPLICANT: Nakamura, Toru

Db 13 EEBKR 18

US-09-834-759-544

Sequence 544, Application US/09834759

Patent No. 6680197

GENERAL INFORMATION:

APPLICANT: Jiang, Yugu

APPLICANT: Dillon, Davin C.

APPLICANT: Mitcham, Jennifer L.

APPLICANT: Xu, Jiangchun

APPLICANT: Harlocker, Susan L.

APPLICANT: Hepler, William T.

APPLICANT: Henderson, Robert A.

TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY AND

FILE REFERENCE: 210121.470C9

CURRENT FILING DATE: 2001-04-13

NUMBER OF SEQ ID NOS: 547

SOFTWARE: FastSeq for Windows Version 3.0

SEQ ID NO 544

LENGTH: 20

TYPE: PRT

ORGANISM: Homo sapiens

US-09-834-759-544

Query Match 1.1%; Score 6; DB 4; Length 20;

Best Local Similarity 100.0%; Pred. No. 2.2e+02; Indels 0; Gaps 0;

Matches 6; Conservative 0; Mismatches 0;

QY 316 LQQQLV 321

Db 5 LQQQLV 10

US-09-834-759-543

Sequence 543, Application US/09834759

Patent No. 6680197

GENERAL INFORMATION:

APPLICANT: Jiang, Yugu

APPLICANT: Dillon, Davin C.

APPLICANT: Mitcham, Jennifer L.

APPLICANT: Xu, Jiangchun

APPLICANT: Harlocker, Susan L.

APPLICANT: Hepler, William T.

APPLICANT: Henderson, Robert A.

TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY AND

FILE REFERENCE: 210121.470C9

CURRENT FILING DATE: 2001-04-13

NUMBER OF SEQ ID NOS: 547

SOFTWARE: FastSeq for Windows Version 3.0

SEQ ID NO 543

LENGTH: 21

TYPE: PRT

ORGANISM: Homo sapiens

US-09-834-759-543

Query Match 1.1%; Score 6; DB 4; Length 21;

Best Local Similarity 100.0%; Pred. No. 2.3e+02; Indels 0; Gaps 0;

Matches 6; Conservative 0; Mismatches 0;

QY 316 LQQQLV 321

Db 14 LQQQLV 19

US-08-851-843A-113

Sequence 233, Application US/08974549A

Patent No. 6166178

GENERAL INFORMATION:

APPLICANT: Cech, Thomas R.

APPLICANT: Lingner, Joachim

APPLICANT: Nakamura, Toru

APPLICANT: Chapman, Karen B.
APPLICANT: Morin, Gregg B.
APPLICANT: Harley, Calvin B.
APPLICANT: Andrews, William H.
TITLE OF INVENTION: Human Telomerase Catalytic Subunit
NUMBER OF SEQUENCES: 727
CORRESPONDENCE ADDRESS:
ADDRESSEE: Townsend and Townsend and Crew LLP
STREET: Two Embarcadero Center, Eighth Floor
CITY: San Francisco
STATE: California
COUNTRY: USA
ZIP: 94111-3634
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/974,549A
FILING DATE: 19-NOV-1997
CLASSIFICATION: 536
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/724,643
FILING DATE: 01-OCT-1996
APPLICATION DATA:
APPLICATION NUMBER: US 08/844,419
FILING DATE: 18-APR-1997
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/846,017
FILING DATE: 25-APR-1997
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/851,843
FILING DATE: 06-MAY-1997
APPLICATION DATA:
APPLICATION NUMBER: US 08/854,050
FILING DATE: 09-MAY-1997
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/911,312
FILING DATE: 14-AUG-1997
APPLICATION DATA:
APPLICATION NUMBER: US 08/912,951
FILING DATE: 14-AUG-1997
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/915,503
FILING DATE: 14-AUG-1997
APPLICATION DATA:
APPLICATION NUMBER: WO PCT/US97/17618
FILING DATE: 01-OCT-1997
PRIOR APPLICATION DATA:
APPLICATION NUMBER: WO PCT/US97/17885
FILING DATE: 01-OCT-1997
ATTORNEY/AGENT INFORMATION:
NAME: Apple, Randolph Ted
REGISTRATION NUMBER: 36,429
REFERENCE/DOCKET NUMBER: 015389-002610US
TELECOMMUNICATION INFORMATION:
TELEPHONE: (415) 576-0300
TELEFAX: (415) 576-0300
INFORMATION FOR SEQ ID NO: 233:
SEQUENCE CHARACTERISTICS:
LENGTH: 23 amino acids
TYPE: amino acid
STRANDEDNESS:
TOPOLOGY: linear
MOLECULE TYPE: peptide
US-08-974-549A-233

Query Match 1.1%; Score 6; DB 3; Length 23;
Best Local Similarity 100.0%; Pred. No. 2.5e+02;
Matches 6; Conservative 0; Mismatches 0; Indels 0;

QY 52 LRDVSE 57

DB 9 LRDVSE 14

RESULT 91
US-08-854-050-113
Sequence 113, Application US/08854050
Patent No. 6261836
GENERAL INFORMATION:
APPLICANT: Cech, Thomas R.
APPLICANT: Lingner, Joachim
APPLICANT: Nakamura, Toru
APPLICANT: Chapman, Karen B.
APPLICANT: Morin, Gregg B.
APPLICANT: Harley, Calvin
APPLICANT: Andrews, William H.
TITLE OF INVENTION: No. 6261836el Telomerase
NUMBER OF SEQUENCES: 225
CORRESPONDENCE ADDRESS:
ADDRESSEE: Townsend and Townsend and Crew LLP
STREET: Two Embarcadero Center, 8th Floor
CITY: San Francisco
STATE: California
COUNTRY: United States of America
ZIP: 94111
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/854,050
FILING DATE: 09-MAY-1997
CLASSIFICATION: 536
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/851,843
FILING DATE: 06-MAY-1997
APPLICATION DATA:
APPLICATION NUMBER: US 08/846,017
FILING DATE: 25-APR-1997
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/844,419
FILING DATE: 18-APR-1997
CLASSIFICATION: 536
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/724,643
FILING DATE: 01-OCT-1996
CLASSIFICATION: 536
ATTORNEY/AGENT INFORMATION:
NAME: Apple, Randolph T.
REGISTRATION NUMBER: 36,429
REFERENCE/DOCKET NUMBER: 015389-002930US
TELECOMMUNICATION INFORMATION:
TELEPHONE: (415) 576-0200
TELEFAX: (415) 576-0300
INFORMATION FOR SEQ ID NO: 113:
SEQUENCE CHARACTERISTICS:
LENGTH: 23 amino acids
TYPE: amino acid
STRANDEDNESS:
TOPOLOGY: linear
MOLECULE TYPE: peptide
US-08-854-050-113

Query Match 1.1%; Score 6; DB 3; Length 23;
Best Local Similarity 100.0%; Pred. No. 2.5e+02;
Matches 6; Conservative 0; Mismatches 0; Indels 0;

QY 52 LRDVSE 57

DB 9 LRDVSE 14

RESULT 92

US-09-430-323-113
; Sequence 113, Application US/09430323
; Patent No. 6309867

GENERAL INFORMATION:

APPLICANT: Cech, Thomas R.
; Lingner, Joachim
; Nakamura, Toru
; Chapman, Karen B.
; Morin, Gregg B.
; Harley, Calvin
; Andrews, William H.
TITLE OF INVENTION: NO. 6309867el Telomerase
NUMBER OF SEQUENCES: 225

ADDRESS: Townsend and Townsend and Crew LLP
STREET: Two Embarcadero Center, 8th Floor
CITY: San Francisco
STATE: California
COUNTRY: United States of America
ZIP: 94111

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/430,323
FILING DATE: 29-Oct-1999
CLASSIFICATION: <Unknown>

PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/854,050
FILING DATE: 09-MAY-1997
CLASSIFICATION: 435
APPLICATION NUMBER: US 08/846,017
FILING DATE: 25-APR-1997
APPLICATION NUMBER: US 08/844,419
FILING DATE: 18-APR-1997
APPLICATION NUMBER: US 08/724,643
FILING DATE: 01-OCT-1996

ATTORNEY/AGENT INFORMATION:
NAME: Apple, Randolph T.
REGISTRATION NUMBER: 36,429
REFERENCE/DOCKET NUMBER: 015389-002930US
TELEPHONE: (415) 576-0200
TELEFAX: (415) 576-0300

INFORMATION FOR SEQ ID NO: 113:
SEQUENCE CHARACTERISTICS:
LENGTH: 23 amino acids
TYPE: amino acid
STRANDEDNESS: <Unknown>
TOPOLOGY: linear
MOLECULE TYPE: peptide
SEQUENCE DESCRIPTION: SEQ ID NO: 113:
US-09-430-323-113

Query Match 1.1%; Score 6; DB 4; Length 23;
Best Local Similarity 100.0%; Pred. No. 2.5e+02;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 52 LRDVSE 57
| | | | |
Db 9 LRDVSE 14

RESULT 93

US-08-912-951-114
; Sequence 114, Application US/08912951
; Patent No. 6475789

GENERAL INFORMATION:

APPLICANT: Cech, Thomas R.
; Lingner, Joachim
; Nakamura, Toru
; Chapman, Karen B.
; Morin, Gregg B.
; Harley, Calvin
; Andrews, William H.
TITLE OF INVENTION: HUMAN TELOMERASE CATALYTIC SUBUNIT: DIAGNOSTIC AND
THERAPEUTIC METHODS
NUMBER OF SEQUENCES: 335
CORRESPONDENCE ADDRESS:
ADDRESSEE: Townsend and Townsend and Crew LLP
STREET: Two Embarcadero Center, 8th Floor
CITY: San Francisco
STATE: California
COUNTRY: United States of America
ZIP: 94111

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/912,951
FILING DATE: 14-AUG-1997
CLASSIFICATION: 435

PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/854,050
FILING DATE: 09-MAY-1997
CLASSIFICATION: 435
APPLICATION NUMBER: US 08/846,017
FILING DATE: 25-APR-1997
CLASSIFICATION: 435

PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/844,419
FILING DATE: 18-APR-1997
CLASSIFICATION: 435
APPLICATION NUMBER: US 08/724,643
FILING DATE: 01-OCT-1996

ATTORNEY/AGENT INFORMATION:
NAME: Apple, Randolph T.
REGISTRATION NUMBER: 36,429
REFERENCE/DOCKET NUMBER: 015389-002600US
TELEPHONE: (415) 576-0200
TELEFAX: (415) 576-0300
INFORMATION FOR SEQ ID NO: 114:
SEQUENCE CHARACTERISTICS:
LENGTH: 23 amino acids
TYPE: amino acid
STRANDEDNESS:
TOPOLOGY: linear
MOLECULE TYPE: peptide
US-08-912-951-114

Query Match 1.1%; Score 6; DB 4; Length 23;
Best Local Similarity 100.0%; Pred. No. 2.5e+02;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 52 LRDVSE 57
| | | | |
Db 9 LRDVSE 14

RESULT 94

US-09-402-181B-233
; Sequence 233, Application US/09402181B
; Patent No. 6610839
; GENERAL INFORMATION:
; APPLICANT: Cech, Thomas R.
; Lingner, Joachim
; Nakamura, Toru
; Chapman, Karen B.
; Morin, Gregg B.
; Harley, Calvin B.
; Andrews, William H.
; TITLE OF INVENTION: Human Telomerase Catalytic Subunit
; NUMBER OF SEQUENCES: 633
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Townsend and Townsend and Crew LLP
; STREET: Two Embarcadero Center, Eighth Floor
; CITY: San Francisco
; STATE: California
; COUNTRY: USA
; ZIP: 94111-3834
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION NUMBER: US/09/402,181B
; FILING DATE: 29-Sep-1997
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/724,643
; FILING DATE: 01-OCT-1996
; APPLICATION NUMBER: US 08/844,419
; FILING DATE: 18-APR-1997
; APPLICATION NUMBER: US 08/846,017
; FILING DATE: 25-APR-1997
; APPLICATION NUMBER: US 08/851,843
; FILING DATE: 06-MAY-1997
; APPLICATION NUMBER: US 08/854,050
; FILING DATE: 09-MAY-1997
; APPLICATION NUMBER: US 08/911,312
; FILING DATE: 14-AUG-1997
; APPLICATION NUMBER: US 08/912,951
; FILING DATE: 14-AUG-1997
; APPLICATION NUMBER: WO PCT/US97/17885
; FILING DATE: 01-OCT-1997
; ATTORNEY/AGENT INFORMATION:
; NAME: Ausubus, Scott L.
; REGISTRATION NUMBER: 42,271
; REFERENCE/DOCKET NUMBER: 015389-002620US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 576-0200
; TELEFAX: (415) 576-0300
; INFORMATION FOR SEQ ID NO: 233:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 23 amino acids
; TYPE: amino acid
; STRANDEDNESS: <Unknown>
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; SEQUENCE DESCRIPTION: SEQ ID NO: 233:
US-09-402-181B-233

Query Match 1.1%; Score 6; DB 4; Length 23;
Best local Similarity 100.0%; Pred. No. 2.5e+02;
Matches 6; Conservative 0; Mismatches 0; Indels 0;
Gaps 0;
52 LRDVSE 57
|||||
9 LRDVSE 14

RESULT 95
US-09-721-456-233
; Sequence 233, Application US/09721456
; Patent No. 6617110
; GENERAL INFORMATION:
; APPLICANT: Cech, Thomas R.
; Lingner, Joachim
; Nakamura, Toru
; Chapman, Karen B.
; Morin, Gregg B.
; Harley, Calvin B.
; Andrews, William H.
; TITLE OF INVENTION: Human Telomerase Catalytic Subunit
; NUMBER OF SEQUENCES: 727
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Townsend and Townsend and Crew LLP
; STREET: Two Embarcadero Center, Eighth Floor
; CITY: San Francisco
; STATE: California
; COUNTRY: USA
; ZIP: 94111-3834
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION NUMBER: US/09/721,456
; FILING DATE: 22-No. 6617110-2000
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/974,549A
; FILING DATE: 19-NOV-1997
; APPLICATION NUMBER: US 08/724,643
; FILING DATE: 01-OCT-1996
; APPLICATION NUMBER: US 08/844,419
; FILING DATE: 18-APR-1997
; APPLICATION NUMBER: US 08/846,017
; FILING DATE: 25-APR-1997
; APPLICATION NUMBER: US 08/851,843
; FILING DATE: 06-MAY-1997
; APPLICATION NUMBER: US 08/854,050
; FILING DATE: 09-MAY-1997
; APPLICATION NUMBER: US 08/911,312
; FILING DATE: 14-AUG-1997
; APPLICATION NUMBER: US 08/912,951
; FILING DATE: 14-AUG-1997
; APPLICATION NUMBER: US 08/915,503
; FILING DATE: 14-AUG-1997
; APPLICATION NUMBER: WO PCT/US97/17618
; FILING DATE: 01-OCT-1997
; APPLICATION NUMBER: WO PCT/US97/17885
; FILING DATE: 01-OCT-1997
; ATTORNEY/AGENT INFORMATION:
; NAME: Apple, Randolph Ted
; REGISTRATION NUMBER: 36,429
; REFERENCE/DOCKET NUMBER: 015389-002610US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 576-0200
; TELEFAX: (415) 576-0300
; INFORMATION FOR SEQ ID NO: 233:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 23 amino acids
; TYPE: amino acid
; STRANDEDNESS: <Unknown>
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; SEQUENCE DESCRIPTION: SEQ ID NO: 233:
US-09-721-456-233

Query Match 1.1%; Score 6; DB 4; Length 23;
Best local Similarity 100.0%; Pred. No. 2.5e+02;

Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 52 LRWSE 57
|||||
DQ 9 LRWSE 14

RESULT 96
US-08-507-598-4
; Sequence 4, Application US/08507598
; Patent No. 5934188
; GENERAL INFORMATION:
; APPLICANT: HARADA, SHUN-ICHI
; APPLICANT: SAMPATH, T. K.
; APPLICANT: RODAN, GIDEON A.
; TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR IDENTIFYING
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: PATENT ADMINISTRATOR, TESTA, HURWITZ &
; ADDRESSEE: THIBEAULT
; STREET: 53 STATE STREET
; CITY: BOSTON
; STATE: MA
; COUNTRY: USA
; ZIP: 02109
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/507,598
; FILING DATE:
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: PITCHER, EDMUND R.
; REGISTRATION NUMBER: 27,829
; REFERENCE/DOCKET NUMBER: CRP-107
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617)-248-7000
; TELEFAX: (617)-248-7100
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 25 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; FEATURE:
; NAME/KEY: Peptide
; LOCATION: 1..25
; OTHER INFORMATION: /note= "Conserved domain of human
; US-08-507-598-4

Query Match 1.1%; Score 6; DB 2; Length 25;
Best Local Similarity 100.0%; Pred. No. 2.7e+02;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 249 EEEKR 254
|||||
DQ 8 EEEKR 13

RESULT 97
US-08-507-750-4
; Sequence 4, Application US/08507750
; Patent No. 5932716
; GENERAL INFORMATION:
; APPLICANT: SAMPATH, T. K.
; TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR IDENTIFYING
; NUMBER OF SEQUENCES: 10
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: PATENT ADMINISTRATOR, TESTA, HURWITZ &
; ADDRESSEE: THIBEAULT
; STREET: 53 STATE STREET
; CITY: BOSTON
; STATE: MA
; COUNTRY: USA
; ZIP: 02109
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/764,522A

NUMBER OF SEQUENCES: 4
CORRESPONDENCE ADDRESS:
ADDRESSEE: PATENT ADMINISTRATOR, TESTA, HURWITZ &
ADDRESSEE: THIBEAULT
STREET: 53 STATE STREET
CITY: BOSTON
STATE: MA
COUNTRY: USA
ZIP: 02109
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/507,750
FILING DATE:
CLASSIFICATION: 424
ATTORNEY/AGENT INFORMATION:
NAME: PITCHER, EDMUND R.
REGISTRATION NUMBER: 27,829
REFERENCE/DOCKET NUMBER: CRP-116
TELECOMMUNICATION INFORMATION:
TELEPHONE: (617)-248-7000
TELEFAX: (617)-248-7100
INFORMATION FOR SEQ ID NO: 4:
SEQUENCE CHARACTERISTICS:
LENGTH: 25 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: peptide
FEATURE:
NAME/KEY: Peptide
LOCATION: 1..25
OTHER INFORMATION: /note= "Conserved domain of human
US-08-507-750-4

Query Match 1.1%; Score 6; DB 2; Length 25;
Best Local Similarity 100.0%; Pred. No. 2.7e+02;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 249 EEEKR 254
|||||
DQ 8 EEEKR 13

RESULT 98
US-08-764-522A-4
; Sequence 4, Application US/08764522A
; Patent No. 6090544
; GENERAL INFORMATION:
; APPLICANT: HARADA, SHUN-ICHI
; APPLICANT: SAMPATH, T. K.
; APPLICANT: RODAN, GIDEON A.
; TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR IDENTIFYING
; TITLE OF INVENTION: MORPHOGEN ANALOGS
; NUMBER OF SEQUENCES: 10
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: PATENT ADMINISTRATOR, CREATIVE BIOMOLECULES
; ADDRESSEE: THIBEAULT
; STREET: 45 SOUTH STREET
; CITY: HOPKINTON
; STATE: MA
; COUNTRY: USA
; ZIP: 01748
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/764,522A

; FILING DATE: 435
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: VITO, CHRISTINE C.
; REGISTRATION NUMBER: 39,061
; REFERENCE/DOCKET NUMBER: CRP-126
; TELEPHONE: (617)-248-7000
; TELEFAX: (617)-248-7100
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 25 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: Peptide
; FEATURE:
; NAME/KEY: Peptide
; LOCATION: 1..25
; OTHER INFORMATION: /note= "Conserved domain of human
; OTHER INFORMATION: c-fos"
; US-08-764-522A-4
; Query Match 1.1%; Score 6; DB 3; Length 25;
; Best Local Similarity 100.0%; Pred. No. 2.7e+02;
; Matches 6; Conservative 0; Mismatches 0; Indels 0;
; Gaps 0;
; QY 249 EEEER 254
; DB 8 EEEER 13
; RESULT 99
; US-08-764-528-4
; Sequence 4, Application US/08764528
; Patent No. 6103491
; GENERAL INFORMATION:
; APPLICANT: SAMPATH, K. T.
; TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR IDENTIFYING
; MORPHOGEN ANALOGS
; NUMBER OF SEQUENCES: 8
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: PATENT ADMINISTRATOR, CREATIVE BIOMOLECULES
; STREET: 45 SOUTH STREET
; CITY: HOPKINTON
; STATE: MA
; COUNTRY: USA
; ZIP: 01748
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; FILING DATE: 11-JUN-1997
; CLASSIFICATION: 424
; ATTORNEY/AGENT INFORMATION:
; NAME: VITO, CHRISTINE C.
; REGISTRATION NUMBER: 39,061
; REFERENCE/DOCKET NUMBER: CRP-127
; TELEPHONE: (617)-248-7000
; TELEFAX: (617)-248-7100
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 25 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; FEATURE:
; NAME/KEY: Peptide
; LOCATION: 1..25
; OTHER INFORMATION: /note= "Conserved domain of human
; OTHER INFORMATION: c-fos"
; US-08-764-522A-4

; LOCATION: 1..25
; OTHER INFORMATION: /note= "Conserved domain of human
; OTHER INFORMATION: c-fos"
; US-08-764-528-4
; Query Match 1.1%; Score 6; DB 3; Length 25;
; Best Local Similarity 100.0%; Pred. No. 2.7e+02;
; Matches 6; Conservative 0; Mismatches 0; Indels 0;
; Gaps 0;
; QY 249 EEEER 254
; DB 8 EEEER 13
; RESULT 100
; US-08-872-859-4
; Sequence 4, Application US/08872859
; Patent No. 6110460
; GENERAL INFORMATION:
; APPLICANT: SAMPATH, T. K.
; TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR IDENTIFYING
; MORPHOGEN ANALOGS
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: PATENT ADMINISTRATOR, TESTA, HURWITZ &
; STREET: 53 STATE STREET
; CITY: BOSTON
; STATE: MA
; COUNTRY: USA
; ZIP: 02109
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; FILING DATE: 11-JUN-1997
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/507,750
; FILING DATE: 26-JUL-1995
; ATTORNEY/AGENT INFORMATION:
; NAME: PITCHER, EDMUND R.
; REGISTRATION NUMBER: 27,829
; REFERENCE/DOCKET NUMBER: CRP-116
; TELEPHONE: (617)-248-7000
; TELEFAX: (617)-248-7100
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 25 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; FEATURE:
; NAME/KEY: Peptide
; LOCATION: 1..25
; OTHER INFORMATION: /note= "Conserved domain of human
; OTHER INFORMATION: c-fos"
; US-08-872-859-4
; Query Match 1.1%; Score 6; DB 3; Length 25;
; Best Local Similarity 100.0%; Pred. No. 2.7e+02;
; Matches 6; Conservative 0; Mismatches 0; Indels 0;
; Gaps 0;
; QY 249 EEEER 254
; DB 8 EEEER 13

Mon Jun 7 17:24:59 2004

us-10-023-529-8.oli.ra1

Page 36

Search completed: June 7, 2004, 14:47:22
Job time : 25 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: June 7, 2004, 14:45:27 ; Search time 41 Seconds
(without alignments)
3636.817 Million cell updates/sec

Title: US-10-023-529-8

Perfect score: 530

Sequence: 1 KSSPPQZAGPCAQRPSQ.....APSTEASGQTGQEPSTARA 530

Scoring table: OLIGO
Gapop 60.0 , Gapext 60.0

Searched: 1155919 seqs, 281338677 residues

Word size : 0

Total number of hits satisfying chosen parameters: 1155919

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 300 summaries

Database : Published Applications AA:
1: /cgn2_6/prodata/1/pubpaa/US07_PUBCOMB.pep.*
2: /cgn2_6/prodata/1/pubpaa/PCT_NEW_PUB.pep.*
3: /cgn2_6/prodata/1/pubpaa/US06_NEW_PUB.pep.*
4: /cgn2_6/prodata/1/pubpaa/US06_PUBCOMB.pep.*
5: /cgn2_6/prodata/1/pubpaa/PCTUS_NEW_PUB.pep.*
6: /cgn2_6/prodata/1/pubpaa/PCTUS_PUBCOMB.pep.*
7: /cgn2_6/prodata/1/pubpaa/US08_NEW_PUB.pep.*
8: /cgn2_6/prodata/1/pubpaa/US08_PUBCOMB.pep.*
9: /cgn2_6/prodata/1/pubpaa/US09A_PUBCOMB.pep.*
10: /cgn2_6/prodata/1/pubpaa/US09B_PUBCOMB.pep.*
11: /cgn2_6/prodata/1/pubpaa/US09C_PUBCOMB.pep.*
12: /cgn2_6/prodata/1/pubpaa/US09_NEW_PUB.pep.*
13: /cgn2_6/prodata/1/pubpaa/US10A_PUBCOMB.pep.*
14: /cgn2_6/prodata/1/pubpaa/US10B_PUBCOMB.pep.*
15: /cgn2_6/prodata/1/pubpaa/US10C_PUBCOMB.pep.*
16: /cgn2_6/prodata/1/pubpaa/US10_NEW_PUB.pep.*
17: /cgn2_6/prodata/1/pubpaa/US60_NEW_PUB.pep.*
18: /cgn2_6/prodata/1/pubpaa/US60_PUBCOMB.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	530	100.0	530	9	US-09-962-055-8
2	530	100.0	530	9	US-09-976-740-8
3	530	100.0	530	12	US-10-671-242-8
4	530	100.0	530	13	US-10-023-529-8
5	530	100.0	530	13	US-10-023-523-8
6	530	100.0	530	15	US-10-616-187-8
7	429	80.9	546	9	US-09-976-740-44
8	429	80.9	546	12	US-10-671-242-44
9	429	80.9	546	13	US-10-023-529-44
10	429	80.9	546	13	US-10-023-523-44
11	429	80.9	546	15	US-10-616-187-44
12	359	67.7	510	12	US-10-276-774-2134
13	172	32.5	557	9	US-09-962-055-5
14	172	32.5	557	9	US-09-976-740-5
15	172	32.5	557	12	US-10-671-242-5

16	172	32.5	557	13	US-10-023-529-5	Sequence 5, Appli
17	172	32.5	557	13	US-10-023-523-5	Sequence 5, Appli
18	172	32.5	557	15	US-10-616-187-5	Sequence 5, Appli
19	17	3.2	85	15	US-10-264-049-3413	Sequence 3413, Ap
20	11	2.1	436	12	US-10-424-599-207434	Sequence 207434,
21	9	1.7	99	9	US-09-764-869-1082	Sequence 1082, Ap
22	23	9	99	14	US-10-091-504-1082	Sequence 1082, Ap
23	9	1.7	99	15	US-10-227-577-1082	Sequence 112, App
24	9	1.7	594	9	US-09-801-368-112	Sequence 112, App
25	9	1.7	594	15	US-10-369-493-1762	Sequence 1762, Ap
26	9	1.7	637	12	US-10-382-122A-72115	Sequence 72115, A
27	8	1.7	637	12	US-10-443-622-130	Sequence 130, App
28	8	1.5	106	12	US-10-425-114-48446	Sequence 48446, A
29	8	1.5	121	12	US-10-425-114-48446	Sequence 272074
30	8	1.5	156	16	US-10-443-622-122	Sequence 122, App
31	8	1.5	229	14	US-10-080-170-316	Sequence 316, App
32	8	1.5	413	14	US-10-156-761-12433	Sequence 12433, A
33	8	1.5	447	15	US-10-369-493-21212	Sequence 21212, A
34	8	1.5	450	12	US-10-335-977-7865	Sequence 7865, Ap
35	8	1.5	451	12	US-10-335-977-7866	Sequence 7866, Ap
36	8	1.5	484	9	US-09-738-626-6430	Sequence 6430, Ap
37	8	1.5	525	15	US-10-389-493-23425	Sequence 23425, A
38	7	1.3	7	9	US-09-962-055-41	Sequence 41, Appl
39	7	1.3	7	9	US-09-976-740-41	Sequence 41, Appl
40	7	1.3	7	12	US-10-671-242-41	Sequence 41, Appl
41	7	1.3	7	13	US-10-023-529-41	Sequence 41, Appl
42	7	1.3	7	13	US-10-023-523-41	Sequence 41, Appl
43	7	1.3	7	13	US-10-616-187-41	Sequence 41, Appl
44	7	1.3	18	14	US-10-225-567A-2071	Sequence 2071, Ap
45	7	1.3	27	14	US-10-195-730-348	Sequence 348, App
46	7	1.3	38	9	US-09-864-761-36055	Sequence 36055, A
47	7	1.3	44	10	US-09-491-614-31	Sequence 31, Appl
48	7	1.3	45	10	US-09-986-480-335	Sequence 335, App
49	7	1.3	45	10	US-09-892-877-355	Sequence 355, App
50	7	1.3	45	10	US-09-948-783-359	Sequence 359, App
51	7	1.3	55	12	US-10-424-599-154654	Sequence 154654,
52	7	1.3	78	12	US-10-424-599-246225	Sequence 246225,
53	7	1.3	80	12	US-10-425-114-68147	Sequence 68147, A
54	7	1.3	82	9	US-09-796-692-1230	Sequence 1230, Ap
55	7	1.3	82	14	US-10-040-862-1230	Sequence 1230, Ap
56	7	1.3	82	15	US-10-057-475B-1230	Sequence 1230, Ap
57	7	1.3	82	15	US-10-154-884B-1230	Sequence 1230, Ap
58	7	1.3	82	15	US-10-057-475B-1236	Sequence 1236, Ap
59	7	1.3	91	9	US-09-796-692-1236	Sequence 1236, Ap
60	7	1.3	91	9	US-09-796-692-1236	Sequence 1236, Ap
61	7	1.3	91	14	US-10-040-862-1236	Sequence 1236, Ap
62	7	1.3	91	14	US-10-040-862-1236	Sequence 1236, Ap
63	7	1.3	91	15	US-10-057-475B-1236	Sequence 1236, Ap
64	7	1.3	91	15	US-10-057-475B-1236	Sequence 1236, Ap
65	7	1.3	91	15	US-10-154-884B-1236	Sequence 1236, Ap
66	7	1.3	91	15	US-10-154-884B-1236	Sequence 1236, Ap
67	7	1.3	95	9	US-09-864-761-39633	Sequence 39633, A
68	7	1.3	99	12	US-09-804-014A-56	Sequence 56, Appl
69	7	1.3	100	12	US-10-424-599-266203	Sequence 266203,
70	7	1.3	101	12	US-10-424-599-153810	Sequence 153810,
71	7	1.3	106	15	US-10-264-049-4293	Sequence 4293, Ap
72	7	1.3	110	14	US-10-195-730-345	Sequence 345, App
73	7	1.3	113	12	US-10-425-114-50441	Sequence 50441, A
74	7	1.3	114	12	US-10-424-599-165582	Sequence 165582,
75	7	1.3	115	14	US-10-424-599-196307	Sequence 196307,
76	7	1.3	115	14	US-10-043-487-334	Sequence 334, App
77	7	1.3	119	12	US-10-335-977-4955	Sequence 4955, Ap
78	7	1.3	119	12	US-10-335-977-4955	Sequence 4955, Ap
79	7	1.3	122	16	US-10-424-599-261987	Sequence 261987,
80	7	1.3	122	16	US-10-443-622-121	Sequence 121, App
81	7	1.3	124	12	US-10-424-599-150351	Sequence 150351,
82	7	1.3	127	8	US-08-849-303-19	Sequence 19, Appl
83	7	1.3	130	15	US-10-108-260A-3437	Sequence 3437, Ap
84	7	1.3	133	12	US-10-335-977-4956	Sequence 4956, Ap
85	7	1.3	133	12	US-10-335-977-4957	Sequence 4957, Ap
86	7	1.3	137	12	US-10-335-977-4958	Sequence 4958, Ap
87	7	1.3	138	10	US-09-847-208-130	Sequence 130, App
88	7	1.3	139	10	US-09-847-208-130	Sequence 128, App
89	7	1.3	142	12	US-10-425-114-69680	Sequence 69680, A

89	7	1.3	146	12	US-10-424-599-167688	Sequence 167688,	162	7	1.3	350	12	US-10-425-114-52194	Sequence 52194, A
90	7	1.3	156	10	US-09-893-519A-11	Sequence 11, Appl	163	7	1.3	350	14	US-10-309-851-8	Sequence 8, Appl
91	7	1.3	159	9	US-09-934-289A-20	Sequence 20, Appl	164	7	1.3	351	12	US-10-283-122A-76372	Sequence 76372, A
92	7	1.3	166	12	US-10-424-599-163595	Sequence 163595,	165	7	1.3	352	12	US-10-424-599-180936	Sequence 180936,
93	7	1.3	167	12	US-10-424-599-241739	Sequence 241739,	166	7	1.3	352	12	US-10-425-114-57918	Sequence 57918, A
94	7	1.3	171	12	US-10-424-599-262163	Sequence 262163,	167	7	1.3	352	12	US-10-425-114-46810	Sequence 46810, A
95	7	1.3	176	10	US-09-847-708-129	Sequence 129, App	168	7	1.3	355	12	US-10-425-114-64921	Sequence 64921, A
96	7	1.3	186	12	US-10-276-774-2467	Sequence 2467, Ap	169	7	1.3	356	12	US-10-425-114-47534	Sequence 47534, A
97	7	1.3	197	9	US-09-934-289A-18	Sequence 18, Appl	170	7	1.3	362	12	US-10-369-493-13566	Sequence 13566, A
98	7	1.3	206	12	US-10-425-114-45419	Sequence 45419, A	171	7	1.3	363	15	US-10-282-122A-54378	Sequence 54378, A
99	7	1.3	207	15	US-10-104-047-2607	Sequence 2607, Ap	172	7	1.3	365	12	US-10-424-599-168724	Sequence 168724, A
100	7	1.3	208	12	US-10-424-599-244841	Sequence 244841,	173	7	1.3	366	14	US-10-156-781-10482	Sequence 10482, A
101	7	1.3	215	10	US-09-866-480-339	Sequence 339, App	174	7	1.3	368	15	US-10-264-237-2400	Sequence 2400, Ap
102	7	1.3	215	10	US-09-892-877-359	Sequence 359, App	175	7	1.3	379	14	US-10-050-882-53	Sequence 53, Appl
103	7	1.3	215	10	US-09-948-783-163	Sequence 363, App	176	7	1.3	380	9	US-09-729-674-128	Sequence 128, App
104	7	1.3	215	12	US-10-425-114-43098	Sequence 43098, A	177	7	1.3	381	12	US-10-412-699B-606	Sequence 606, App
105	7	1.3	220	9	US-09-938-803-4	Sequence 4, Appl	178	7	1.3	381	14	US-10-295-403-112	Sequence 112, App
106	7	1.3	228	12	US-10-282-122A-48054	Sequence 48054, A	179	7	1.3	386	15	US-10-004-378A-4	Sequence 4, Appl
107	7	1.3	233	12	US-10-425-114-56502	Sequence 56502, A	180	7	1.3	386	15	US-10-369-493-7832	Sequence 7832, Ap
108	7	1.3	238	12	US-10-282-122A-48996	Sequence 48996, A	181	7	1.3	390	12	US-10-282-122A-65259	Sequence 65259, A
109	7	1.3	240	12	US-10-282-122A-43308	Sequence 43308, A	182	7	1.3	390	12	US-10-282-122A-65956	Sequence 65956, A
110	7	1.3	240	14	US-10-309-851-4	Sequence 4, Appl	183	7	1.3	401	12	US-10-425-114-37468	Sequence 37468, A
111	7	1.3	240	14	US-10-309-851-36	Sequence 36, Appl	184	7	1.3	403	15	US-10-369-493-10660	Sequence 10660, A
112	7	1.3	248	12	US-10-016-248-54	Sequence 54, Appl	185	7	1.3	406	12	US-10-335-977-7931	Sequence 7931, Ap
113	7	1.3	251	12	US-10-412-699B-536	Sequence 536, App	186	7	1.3	416	9	US-09-925-300-1395	Sequence 1395, Ap
114	7	1.3	251	14	US-10-295-403-142	Sequence 142, App	187	7	1.3	418	12	US-10-282-122A-49150	Sequence 49150, A
115	7	1.3	253	12	US-10-424-599-244838	Sequence 244838,	188	7	1.3	418	12	US-10-424-599-153808	Sequence 153808,
116	7	1.3	253	15	US-10-369-493-22886	Sequence 22886, A	189	7	1.3	420	12	US-10-424-599-285648	Sequence 285648,
117	7	1.3	258	12	US-10-016-248-53	Sequence 53, Appl	190	7	1.3	420	12	US-10-425-114-36713	Sequence 36713, A
118	7	1.3	263	12	US-10-425-114-47632	Sequence 47632, A	191	7	1.3	436	15	US-10-104-047-3432	Sequence 3432, Ap
119	7	1.3	264	12	US-10-282-122A-59293	Sequence 59293, A	192	7	1.3	437	12	US-10-425-114-55188	Sequence 55188, A
120	7	1.3	265	14	US-10-309-851-6	Sequence 6, Appl	193	7	1.3	446	9	US-09-824-735-2	Sequence 2, Appl
121	7	1.3	266	12	US-10-380-703-7	Sequence 7, Appl	194	7	1.3	446	12	US-10-617-038-8	Sequence 8, Appl
122	7	1.3	270	14	US-10-156-761-11866	Sequence 11866, A	195	7	1.3	451	15	US-10-004-378A-2	Sequence 2, Appl
123	7	1.3	271	12	US-10-424-599-199397	Sequence 199397,	196	7	1.3	451	12	US-10-282-122A-52293	Sequence 52293, A
124	7	1.3	274	12	US-10-424-599-224138	Sequence 224138,	197	7	1.3	453	12	US-09-978-295A-430	Sequence 430, App
125	7	1.3	275	12	US-10-282-122A-55720	Sequence 55720, A	198	7	1.3	455	9	US-09-978-697-430	Sequence 430, App
126	7	1.3	276	15	US-10-104-047-2279	Sequence 2279, Ap	199	7	1.3	455	9	US-09-978-192A-430	Sequence 430, App
127	7	1.3	277	12	US-10-282-122A-59764	Sequence 59764, A	200	7	1.3	455	9	US-09-965-529-14	Sequence 14, Appl
128	7	1.3	277	14	US-10-106-698-5823	Sequence 5823, Ap	201	7	1.3	455	9	US-09-999-832A-430	Sequence 430, App
129	7	1.3	280	12	US-10-282-122A-78227	Sequence 78227, A	202	7	1.3	455	10	US-09-978-183-430	Sequence 430, App
130	7	1.3	281	12	US-10-335-977-8435	Sequence 8435, Ap	203	7	1.3	455	10	US-09-978-608A-430	Sequence 430, App
131	7	1.3	282	12	US-10-425-114-46688	Sequence 46688, A	204	7	1.3	455	10	US-09-978-585A-430	Sequence 430, App
132	7	1.3	284	12	US-10-424-599-279326	Sequence 279326,	205	7	1.3	455	10	US-09-978-191A-430	Sequence 430, App
133	7	1.3	286	12	US-10-424-599-224134	Sequence 224134,	206	7	1.3	455	10	US-09-978-403A-430	Sequence 430, App
134	7	1.3	292	14	US-10-050-882-78	Sequence 78, Appl	207	7	1.3	455	10	US-09-978-564A-430	Sequence 430, App
135	7	1.3	293	14	US-10-017-161-1866	Sequence 1866, Ap	208	7	1.3	455	10	US-09-999-830A-430	Sequence 430, App
136	7	1.3	293	15	US-10-292-798-1522	Sequence 1522, Ap	209	7	1.3	455	10	US-09-978-824-430	Sequence 430, App
137	7	1.3	296	12	US-10-182-936A-102	Sequence 102, App	210	7	1.3	455	10	US-09-978-757A-430	Sequence 430, App
138	7	1.3	296	15	US-10-374-979-102	Sequence 102, App	211	7	1.3	455	10	US-09-978-187B-430	Sequence 430, App
139	7	1.3	304	15	US-10-369-493-11313	Sequence 11313, A	212	7	1.3	455	10	US-09-978-643A-430	Sequence 430, App
140	7	1.3	306	15	US-09-729-674-70	Sequence 70, Appl	213	7	1.3	455	10	US-09-978-193A-430	Sequence 430, App
141	7	1.3	309	9	US-10-369-493-4700	Sequence 4700, Ap	214	7	1.3	455	10	US-09-978-824-430	Sequence 430, App
142	7	1.3	310	15	US-10-104-047-3795	Sequence 3795, Ap	215	7	1.3	455	10	US-09-969-680A-14	Sequence 14, Appl
143	7	1.3	311	15	US-10-108-260A-2887	Sequence 2887, Ap	216	7	1.3	455	10	US-09-978-375A-430	Sequence 430, App
144	7	1.3	314	12	US-10-425-114-43684	Sequence 43684, A	217	7	1.3	455	10	US-09-978-298A-430	Sequence 430, App
145	7	1.3	316	12	US-10-424-599-256641	Sequence 256641,	218	7	1.3	455	10	US-09-978-188A-430	Sequence 430, App
146	7	1.3	317	12	US-10-282-122A-53118	Sequence 53118, A	219	7	1.3	455	10	US-09-978-188A-430	Sequence 430, App
147	7	1.3	323	12	US-10-425-114-42699	Sequence 42699, A	220	7	1.3	455	10	US-09-978-665A-430	Sequence 430, App
148	7	1.3	326	9	US-09-815-242-11485	Sequence 11485, A	221	7	1.3	455	10	US-09-978-802A-430	Sequence 430, App
149	7	1.3	326	10	US-09-882-227-502	Sequence 227-502	222	7	1.3	455	10	US-10-164-749A-430	Sequence 430, App
150	7	1.3	326	12	US-10-282-122A-58652	Sequence 58652, A	223	7	1.3	455	10	US-10-206-915-14	Sequence 14, Appl
151	7	1.3	327	12	US-10-282-122A-76788	Sequence 76788, A	224	7	1.3	455	10	US-10-199-670-14	Sequence 14, Appl
152	7	1.3	334	12	US-10-425-114-56973	Sequence 56973, A	225	7	1.3	455	10	US-10-201-858-14	Sequence 14, Appl
153	7	1.3	337	12	US-10-424-599-271324	Sequence 271324,	226	7	1.3	455	12	US-09-999-831A-430	Sequence 430, App
154	7	1.3	337	14	US-10-097-111-395	Sequence 111-395	227	7	1.3	455	12	US-10-205-890-14	Sequence 14, Appl
155	7	1.3	337	12	US-10-282-122A-51710	Sequence 51710, A	228	7	1.3	455	12		
156	7	1.3	339	12	US-10-424-599-250813	Sequence 250813,	229	7	1.3	455	12		
157	7	1.3	341	12	US-10-424-599-250815	Sequence 250815,	230	7	1.3	455	12		
158	7	1.3	341	14	US-10-092-066B-1	Sequence 1, Appl	231	7	1.3	455	12		
159	7	1.3	343	12	US-10-424-599-266487	Sequence 266487,	232	7	1.3	455	12		
160	7	1.3	343	12	US-10-424-599-180925	Sequence 180925,	233	7	1.3	455	12		
161	7	1.3	350	12			234	7	1.3	455	12		

7	1.3	455	12	US-10-208-024-14	Sequence 14, Appl
235	7	455	12	US-10-201-853-14	Sequence 14, Appl
236	7	455	12	US-10-063-745-4	Sequence 4, Appl
237	7	455	12	US-10-013-917A-430	Sequence 430, App
238	7	455	12	US-10-063-512-4	Sequence 4, Appl
239	7	455	12	US-10-063-513-4	Sequence 4, Appl
240	7	455	12	US-10-063-515-4	Sequence 4, Appl
241	7	455	12	US-10-063-549-4	Sequence 4, Appl
242	7	455	12	US-10-063-569-4	Sequence 4, Appl
243	7	455	12	US-10-063-551-4	Sequence 4, Appl
244	7	455	12	US-10-174-581-14	Sequence 14, Appl
245	7	455	12	US-10-176-483-14	Sequence 14, Appl
246	7	455	12	US-10-176-749-14	Sequence 14, Appl
247	7	455	12	US-10-176-914-14	Sequence 14, Appl
248	7	455	12	US-10-176-915-14	Sequence 14, Appl
249	7	455	12	US-09-999-834A-430	Sequence 430, App
250	7	455	12	US-10-063-555-4	Sequence 4, Appl
251	7	455	12	US-10-063-563-4	Sequence 4, Appl
252	7	455	12	US-10-063-594-4	Sequence 4, Appl
253	7	455	12	US-10-063-553-4	Sequence 4, Appl
254	7	455	12	US-10-063-554-4	Sequence 4, Appl
255	7	455	12	US-10-176-484-14	Sequence 14, Appl
256	7	455	12	US-10-180-550-14	Sequence 14, Appl
257	7	455	12	US-10-183-014-14	Sequence 14, Appl
258	7	455	12	US-10-187-738-14	Sequence 14, Appl
259	7	455	12	US-10-187-740-14	Sequence 14, Appl
260	7	455	12	US-10-187-883-14	Sequence 14, Appl
261	7	455	12	US-10-194-363-14	Sequence 14, Appl
262	7	455	12	US-10-194-460-14	Sequence 14, Appl
263	7	455	12	US-10-194-463-14	Sequence 14, Appl
264	7	455	12	US-10-195-884-14	Sequence 14, Appl
265	7	455	12	US-10-195-896-14	Sequence 14, Appl
266	7	455	12	US-10-196-744-14	Sequence 14, Appl
267	7	455	12	US-10-196-755-14	Sequence 14, Appl
268	7	455	12	US-10-197-704-14	Sequence 14, Appl
269	7	455	12	US-10-198-758-14	Sequence 14, Appl
270	7	455	12	US-10-198-766-14	Sequence 14, Appl
271	7	455	12	US-10-201-329-14	Sequence 14, Appl
272	7	455	12	US-10-202-412-14	Sequence 14, Appl
273	7	455	12	US-10-206-922-14	Sequence 14, Appl
274	7	455	12	US-10-206-924-14	Sequence 14, Appl
275	7	455	12	US-10-207-921-14	Sequence 14, Appl
276	7	455	12	US-10-207-922-14	Sequence 14, Appl
277	7	455	12	US-10-207-923-14	Sequence 14, Appl
278	7	455	12	US-10-207-924-14	Sequence 14, Appl
279	7	455	12	US-10-208-027-14	Sequence 14, Appl
280	7	455	12	US-10-162-521A-430	Sequence 430, App
281	7	455	12	US-10-174-570-14	Sequence 14, Appl
282	7	455	12	US-10-183-005-14	Sequence 14, Appl
283	7	455	12	US-10-145-016A-430	Sequence 430, App
284	7	455	12	US-10-145-088A-430	Sequence 430, App
285	7	455	12	US-10-145-092A-430	Sequence 430, App
286	7	455	12	US-10-145-129A-430	Sequence 430, App
287	7	455	12	US-10-165-038A-430	Sequence 430, App
288	7	455	12	US-10-165-353A-430	Sequence 430, App
289	7	455	12	US-10-167-600-430	Sequence 430, App
290	7	455	12	US-10-170-481A-430	Sequence 430, App
291	7	455	12	US-10-172-039A-430	Sequence 430, App
292	7	455	12		
293	7	455	12		
294	7	455	12		
295	7	455	12		
296	7	455	12		
297	7	455	12		
298	7	455	12		
299	7	455	12		
300	7	455	12		

ALIGNMENTS

```

1 Sequence 8, Application US/09962055
2 Patent No. US20020052033A1
3 GENERAL INFORMATION:
4 APPLICANT: Lees, Ann M.
5 Lees, Robert S.
6 Law, Simon W.
7 Arjona, Anibal A.
8 TITLE OF INVENTION: NOVEL LOW DENSITY LIPOPROTEIN
9 BINDING PROTEINS AND THEIR USES IN DIAGNOSING AND
10 TREATING ATHEROSCLEROSIS
11
12 NUMBER OF SEQUENCES: 42
13 CORRESPONDENCE ADDRESS:
14 ADDRESSEE: Fish & Richardson P.C.
15 STREET: 225 Franklin Street
16 CITY: Boston
17 STATE: MA
18 COUNTRY: USA
19 ZIP: 02110-2804
20 COMPUTER READABLE FORM:
21 MEDIUM TYPE: Diskette
22 COMPUTER: IBM Compatible
23 OPERATING SYSTEM: DOS
24 SOFTWARE: FastSeq for Windows Version 2.0
25 CURRENT APPLICATION DATA:
26 APPLICATION NUMBER: US/09/962,055
27 FILING DATE: 24-Sep-2001
28 PRIOR APPLICATION DATA:
29 APPLICATION NUMBER: 08/979,608
30 FILING DATE: 26-NOV-1997
31 APPLICATION NUMBER: 50/031,930
32 FILING DATE: 27-NOV-1996
33 ATTORNEY/AGENT INFORMATION:
34 NAME: Myers, Louis
35 REGISTRATION NUMBER: 35,965
36 REFERENCE/DOCKET NUMBER: 10797-002001 (formerly 3983/59818)
37 TELECOMMUNICATION INFORMATION:
38 TELEPHONE: 617/542-5070
39 TELEFAX: 617/542-8906
40 INFORMATION FOR SEQ ID NO: 8:
41 SEQUENCE CHARACTERISTICS:
42 LENGTH: 530 amino acids
43 TYPE: amino acid
44 TOPOLOGY: linear
45 MOLECULE TYPE: protein
46 SEQUENCE DESCRIPTION: SEQ ID NO: 8:
47
48 US-09-962-055-8

```


Db 301 ELREEHIDKVFHKDIQQQLVDAKLQQAQEMLEAEERHQRKDFLLKEAVESQRMCELM 360
Qy 361 KQETHLKOALALYTERFEPONTLSKSSEVFTTFKQEMEMTKKIKKLEKETMYRSRW 420
Db 361 KQETHLKOALALYTERFEPONTLSKSSEVFTTFKQEMEMTKKIKKLEKETMYRSRW 420
Qy 421 ESSNKALLEMAEETKVRDKELEGLQVKIQRLKLCRALQTERNDLNKRVQDLSAGGQSSL 480
Db 421 ESSNKALLEMAEETKVRDKELEGLQVKIQRLKLCRALQTERNDLNKRVQDLSAGGQSSL 480
Qy 481 TDSGPRRPEGGAQAPSSPRVTEAPCYPGAPSTEASGOTGPQPTTSARA 530
Db 481 TDSGPRRPEGGAQAPSSPRVTEAPCYPGAPSTEASGOTGPQPTTSARA 530

RESULT 4
US-10-023-529-8
; Sequence 8, Application US/10023529
; Publication No. US20020129388A1
; GENERAL INFORMATION:
; APPLICANT: Lees, Ann M.
; APPLICANT: Lees, Robert S.
; APPLICANT: Law, Simon W.
; APPLICANT: Arjona, Anibal A.
; TITLE OF INVENTION: NOVEL LOW DENSITY LIPOPROTEIN BINDING
; TITLE OF INVENTION: PROTEINS AND THEIR USE IN DIAGNOSING AND TREATING
; TITLE OF INVENTION: ATHEROSCLEROSIS
; FILE REFERENCE: 10797-004001
; CURRENT APPLICATION NUMBER: US/10/023,529
; CURRENT FILING DATE: 2001-12-17
; PRIOR APPLICATION NUMBER: 09/616,289
; PRIOR FILING DATE: 2000-07-14
; PRIOR APPLICATION NUMBER: US 09/517,849
; PRIOR FILING DATE: 2000-03-02
; PRIOR APPLICATION NUMBER: US 08/979,608
; PRIOR FILING DATE: 1997-11-26
; PRIOR APPLICATION NUMBER: US 60/031,930
; PRIOR FILING DATE: 1996-11-27
; PRIOR APPLICATION NUMBER: US 60/048,547
; PRIOR FILING DATE: 1997-06-03
; NUMBER OF SEQ ID NOS: 53
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 8
; LENGTH: 530
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-023-529-8

Query Match 100.0%; Score 530; DB 13; Length 530;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 530; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 KSSPGQPEAGPEGAQRRPSQAAPAVEAEGPGSSQAAPRKPGEQAQARTAQSGALRDVSELS 60
Db 1 KSSPGQPEAGPEGAQRRPSQAAPAVEAEGPGSSQAAPRKPGEQAQARTAQSGALRDVSELS 60
Qy 61 RQLEDILSTYCVDNNGGPGEDGAGPAEPEDAESKRTYVARNGEPEPTPVVYGEKPS 120
Db 61 RQLEDILSTYCVDNNGGPGEDGAGPAEPEDAESKRTYVARNGEPEPTPVVYGEKPS 120
Qy 121 KGDPTNTEIRQSDVGDGRDHRHPQEKKAAGLGEKITLLMOTLNTLSTPEEKLAALCKKY 180
Db 121 KGDPTNTEIRQSDVGDGRDHRHPQEKKAAGLGEKITLLMOTLNTLSTPEEKLAALCKKY 180
Qy 181 AELLBEHRNSOKMKLLQKKOSQLVQEKDHLRGHSHKAVLARSKLESICRELQHRNLSK 240
Db 181 AELLBEHRNSOKMKLLQKKOSQLVQEKDHLRGHSHKAVLARSKLESICRELQHRNLSK 240
Qy 241 EGVQARAEERBEERKEVTSHPQVTLNDIQLQMEQHNRNSKLRQENNELAERLKKLIEQY 300
Db 241 EGVQARAEERBEERKEVTSHPQVTLNDIQLQMEQHNRNSKLRQENNELAERLKKLIEQY 300
Qy 301 ELREEHIDKVFHKDIQQQLVDAKLQQAQEMLEAEERHQRKDFLLKEAVESQRMCELM 360

Db 301 ELREEHIDKVFHKDIQQQLVDAKLQQAQEMLEAEERHQRKDFLLKEAVESQRMCELM 360
Qy 361 KQETHLKOALALYTERFEPONTLSKSSEVFTTFKQEMEMTKKIKKLEKETMYRSRW 420
Db 361 KQETHLKOALALYTERFEPONTLSKSSEVFTTFKQEMEMTKKIKKLEKETMYRSRW 420
Qy 421 ESSNKALLEMAEETKVRDKELEGLQVKIQRLKLCRALQTERNDLNKRVQDLSAGGQSSL 480
Db 421 ESSNKALLEMAEETKVRDKELEGLQVKIQRLKLCRALQTERNDLNKRVQDLSAGGQSSL 480
Qy 481 TDSGPRRPEGGAQAPSSPRVTEAPCYPGAPSTEASGOTGPQPTTSARA 530
Db 481 TDSGPRRPEGGAQAPSSPRVTEAPCYPGAPSTEASGOTGPQPTTSARA 530

RESULT 5
US-10-023-523-8
; Sequence 8, Application US/10023523
; Publication No. US20020152485A1
; GENERAL INFORMATION:
; APPLICANT: Lees, Ann M.
; APPLICANT: Lees, Robert S.
; APPLICANT: Law, Simon W.
; APPLICANT: Arjona, Anibal A.
; TITLE OF INVENTION: NOVEL LOW DENSITY LIPOPROTEIN BINDING
; TITLE OF INVENTION: PROTEINS AND THEIR USE IN DIAGNOSING AND TREATING
; TITLE OF INVENTION: ATHEROSCLEROSIS
; FILE REFERENCE: 10797-004001
; CURRENT APPLICATION NUMBER: US/10/023,523
; CURRENT FILING DATE: 2001-12-17
; PRIOR APPLICATION NUMBER: US/09/616,289
; PRIOR FILING DATE: 2000-07-14
; PRIOR APPLICATION NUMBER: US 09/517,849
; PRIOR FILING DATE: 2000-03-02
; PRIOR APPLICATION NUMBER: US 08/979,608
; PRIOR FILING DATE: 1997-11-26
; PRIOR APPLICATION NUMBER: US 60/031,930
; PRIOR FILING DATE: 1996-11-27
; PRIOR APPLICATION NUMBER: US 60/048,547
; PRIOR FILING DATE: 1997-06-03
; NUMBER OF SEQ ID NOS: 53
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 8
; LENGTH: 530
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-023-523-8

Query Match 100.0%; Score 530; DB 13; Length 530;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 530; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 KSSPGQPEAGPEGAQRRPSQAAPAVEAEGPGSSQAAPRKPGEQAQARTAQSGALRDVSELS 60
Db 1 KSSPGQPEAGPEGAQRRPSQAAPAVEAEGPGSSQAAPRKPGEQAQARTAQSGALRDVSELS 60
Qy 61 RQLEDILSTYCVDNNGGPGEDGAGPAEPEDAESKRTYVARNGEPEPTPVVYGEKPS 120
Db 61 RQLEDILSTYCVDNNGGPGEDGAGPAEPEDAESKRTYVARNGEPEPTPVVYGEKPS 120
Qy 121 KGDPTNTEIRQSDVGDGRDHRHPQEKKAAGLGEKITLLMOTLNTLSTPEEKLAALCKKY 180
Db 121 KGDPTNTEIRQSDVGDGRDHRHPQEKKAAGLGEKITLLMOTLNTLSTPEEKLAALCKKY 180
Qy 181 AELLBEHRNSOKMKLLQKKOSQLVQEKDHLRGHSHKAVLARSKLESICRELQHRNLSK 240
Db 181 AELLBEHRNSOKMKLLQKKOSQLVQEKDHLRGHSHKAVLARSKLESICRELQHRNLSK 240
Qy 241 EGVQARAEERBEERKEVTSHPQVTLNDIQLQMEQHNRNSKLRQENNELAERLKKLIEQY 300
Db 241 EGVQARAEERBEERKEVTSHPQVTLNDIQLQMEQHNRNSKLRQENNELAERLKKLIEQY 300

Qy 301 ELREEHIDKVPKHKDLQOOLVDKALQQAQEMLKEABERHOREKDFLLKEAVESORMCELM 360
Db 301 ELREEHIDKVPKHKDLQOOLVDKALQQAQEMLKEABERHOREKDFLLKEAVESORMCELM 360
Qy 361 KQOETHLKQOALALYTEKEFEFQNTLSKSSEVFTTFKQEMERMTKKIKLEKETTWYRSRW 420
Db 361 KQOETHLKQOALALYTEKEFEFQNTLSKSSEVFTTFKQEMERMTKKIKLEKETTWYRSRW 420
Qy 421 ESSNKALLEMAEBKTVROKLEGLQVKIQRLEKLCRALQTERNDLNKRVQDLSAGGQSL 480
Db 421 ESSNKALLEMAEBKTVROKLEGLQVKIQRLEKLCRALQTERNDLNKRVQDLSAGGQSL 480
Qy 481 TDSGPERPEGPGCAQAPSSPRVTEAPCYPGAPSTEASGQTGPQEPPTSARA 530
Db 481 TDSGPERPEGPGCAQAPSSPRVTEAPCYPGAPSTEASGQTGPQEPPTSARA 530

RESULT 6

US-10-616-187-8
; Sequence 8, Application US/10616187
; Publication No. US20040013668A1
; GENERAL INFORMATION:
; APPLICANT: Lees, Robert S.
; APPLICANT: Lees, Robert S.
; APPLICANT: Law, Simon W.
; APPLICANT: Arjona, Anibal A.
; TITLE OF INVENTION: NOVEL LOW DENSITY LIPOPROTEIN BINDING
; TITLE OF INVENTION: PROTEINS AND THEIR USE IN DIAGNOSING AND TREATING
; TITLE OF INVENTION: ATHEROSCLEROSIS
; FILE REFERENCE: 10797-004001
; CURRENT APPLICATION NUMBER: US/10/616,187
; CURRENT FILING DATE: 2003-07-09
; PRIOR APPLICATION NUMBER: US/09/616,289
; PRIOR FILING DATE: 2000-07-14
; PRIOR APPLICATION NUMBER: US 09/517,849
; PRIOR FILING DATE: 2000-03-02
; PRIOR APPLICATION NUMBER: US 08/979,608
; PRIOR FILING DATE: 1997-11-26
; PRIOR APPLICATION NUMBER: US 60/031,930
; PRIOR FILING DATE: 1996-11-27
; PRIOR APPLICATION NUMBER: US 60/048,547
; PRIOR FILING DATE: 1997-06-03
; NUMBER OF SEQ ID NOS: 53
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 8
; LENGTH: 530
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-616-187-8

Query Match 100.0%; Score 530; DB 15; Length 530;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 530; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 1 KSSPGQPEAGPGCAQERPSQAAPAVEAEGPGSSQAAPRKEGAQARTAQSGALRDVSEELS 60
Db 1 KSSPGQPEAGPGCAQERPSQAAPAVEAEGPGSSQAAPRKEGAQARTAQSGALRDVSEELS 60
Qy 61 ROLEDILSTVCYDNNQGGGEGDGAQCEPAEPDAEKSRITYVARNGEPEPTPVYVGEKPS 120
Db 61 ROLEDILSTVCYDNNQGGGEGDGAQCEPAEPDAEKSRITYVARNGEPEPTPVYVGEKPS 120
Qy 121 KGDPNTEEIRQSDVEGDRDHRPQEKKAAGLGKEITLLMQTLNTLSTPEEKLAALCKKY 180
Db 121 KGDPNTEEIRQSDVEGDRDHRPQEKKAAGLGKEITLLMQTLNTLSTPEEKLAALCKKY 180
Qy 181 AELLBEHRNSQKMKLLQKQSQLVQEKDHLRGEHSKAVLARSKLESICRELQRNRSK 240
Db 181 AELLBEHRNSQKMKLLQKQSQLVQEKDHLRGEHSKAVLARSKLESICRELQRNRSK 240
Qy 241 EGVQARAEEREEKREKVTSHFQVTLNDIQLQVQEHNRNSKLRQENMELAEKLEKLEIQY 300
Db 241 EGVQARAEEREEKREKVTSHFQVTLNDIQLQVQEHNRNSKLRQENMELAEKLEKLEIQY 300

Qy 301 ELREEHIDKVPKHKDLQOOLVDKALQQAQEMLKEABERHOREKDFLLKEAVESORMCELM 360
Db 301 ELREEHIDKVPKHKDLQOOLVDKALQQAQEMLKEABERHOREKDFLLKEAVESORMCELM 360
Qy 361 KQOETHLKQOALALYTEKEFEFQNTLSKSSEVFTTFKQEMERMTKKIKLEKETTWYRSRW 420
Db 361 KQOETHLKQOALALYTEKEFEFQNTLSKSSEVFTTFKQEMERMTKKIKLEKETTWYRSRW 420
Qy 421 ESSNKALLEMAEBKTVROKLEGLQVKIQRLEKLCRALQTERNDLNKRVQDLSAGGQSL 480
Db 421 ESSNKALLEMAEBKTVROKLEGLQVKIQRLEKLCRALQTERNDLNKRVQDLSAGGQSL 480
Qy 481 TDSGPERPEGPGCAQAPSSPRVTEAPCYPGAPSTEASGQTGPQEPPTSARA 530
Db 481 TDSGPERPEGPGCAQAPSSPRVTEAPCYPGAPSTEASGQTGPQEPPTSARA 530

RESULT 7

US-09-976-740-44
; Sequence 44, Application US/09976740
; Publication No. US20020194633A1
; GENERAL INFORMATION:
; APPLICANT: Lees, Ann M.
; APPLICANT: Lees, Robert S.
; APPLICANT: Law, Simon W.
; APPLICANT: Arjona, Anibal A.
; TITLE OF INVENTION: NOVEL LOW DENSITY LIPOPROTEIN BINDING
; TITLE OF INVENTION: PROTEINS AND THEIR USE IN DIAGNOSING AND TREATING
; TITLE OF INVENTION: ATHEROSCLEROSIS
; FILE REFERENCE: 10797-004001
; CURRENT APPLICATION NUMBER: US/09/976,740
; CURRENT FILING DATE: 2001-10-12
; PRIOR APPLICATION NUMBER: 09/616,289
; PRIOR FILING DATE: 2000-07-14
; PRIOR APPLICATION NUMBER: US 08/979,608
; PRIOR FILING DATE: 1997-11-26
; PRIOR APPLICATION NUMBER: US 60/031,930
; PRIOR FILING DATE: 1996-11-27
; PRIOR APPLICATION NUMBER: US 60/048,547
; PRIOR FILING DATE: 1997-06-03
; NUMBER OF SEQ ID NOS: 53
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 44
; LENGTH: 546
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-976-740-44

Query Match 80.9%; Score 429; DB 9; Length 546;
Best Local Similarity 99.8%; Pred. No. 0;
Matches 529; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
Qy 1 KSSPGQPEAGPGCAQERPSQAAPAVEAEGPGSSQAAPRKEGAQARTAQSGALRDVSEELS 60
Db 17 KSSPGQPEAGPGCAQERPSQAAPAVEAEGPGSSQAAPRKEGAQARTAQSGALRDVSEELS 76
Qy 61 ROLEDILSTVCYDNNQGGGEGDGAQCEPAEPDAEKSRITYVARNGEPEPTPVYVGEKPS 120
Db 77 ROLEDILSTVCYDNNQGGGEGDGAQCEPAEPDAEKSRITYVARNGEPEPTPVYVGEKPS 136
Qy 121 KGDPNTEEIRQSDVEGDRDHRPQEKKAAGLGKEITLLMQTLNTLSTPEEKLAALCKKY 180
Db 137 KGDPNTEEIRQSDVEGDRDHRPQEKKAAGLGKEITLLMQTLNTLSTPEEKLAALCKKY 196
Qy 181 AELLBEHRNSQKMKLLQKQSQLVQEKDHLRGEHSKAVLARSKLESICRELQRNRSK 240
Db 197 AELLBEHRNSQKMKLLQKQSQLVQEKDHLRGEHSKAVLARSKLESICRELQRNRSK 256
Qy 241 EGVQARAEEREEKREKVTSHFQVTLNDIQLQVQEHNRNSKLRQENMELAEKLEKLEIQY 300
Db 257 EGVQARAEEREEKREKVTSHFQVTLNDIQLQVQEHNRNSKLRQENMELAEKLEKLEIQY 316


```

QY 301 ELREHIDKVFHKHLOQQOLVDKLOQAQEMLKEABERHOREKDFLLKEAVESQRMCELM 360
DB 317 ELREHIDKVFHKHLOQQOLVDKLOQAQEMLKEABERHOREKDFLLKEAVESQRMCELM 376
QY 361 KQETHLKQOLALYTEKPEEFQNTLSKSSSEVFTTFKQEMEKMTKKIKLEKETMYRSRW 420
DB 377 KQETHLKQOLALYTEKPEEFQNTLSKSSSEVFTTFKQEMEKMTKKIKLEKETMYRSRW 436
QY 421 ESSNKALLEMABEKTVRDKEGLEQVKIORLEKLCALQTERNDLNKRVODLSAGQGSLS 480
DB 437 ESSNKALLEMABEKTVRDKEGLEQVKIORLEKLCALQTERNDLNKRVODLSAGQGSLS 496
QY 481 TDSGPERRPEGGAQAPSSPRVTEAPCYPGAPSTEASGOTGPQEPSARA 530
DB 497 TDSGPERRPEGGAQAPSSPRVTEAPCYPGAPSTEASGOTGPQEPSARA 546

```

```

RESULT 8
US-10-671-242-44
; Sequence 44, Application US/10671242
; Publication No. US2004040049A1
; GENERAL INFORMATION:
; APPLICANT: Lees, Ann M.
; APPLICANT: Lees, Robert S.
; APPLICANT: Law, Simon W.
; APPLICANT: Arjona, Anibal A.
; TITLE OF INVENTION: NOVEL LOW DENSITY LIPOPROTEIN BINDING
; TITLE OF INVENTION: PROTEINS AND THEIR USE IN DIAGNOSING AND TREATING
; TITLE OF INVENTION: ATHEROSCLEROSIS
; FILE REFERENCE: 10797-004001
; CURRENT APPLICATION NUMBER: US/10/671,242
; CURRENT FILING DATE: 2003-09-24
; PRIOR APPLICATION NUMBER: US/09/616,289
; PRIOR FILING DATE: 2000-07-14
; PRIOR APPLICATION NUMBER: US 09/517,849
; PRIOR FILING DATE: 2000-03-02
; PRIOR APPLICATION NUMBER: US 08/979,608
; PRIOR FILING DATE: 1997-11-26
; PRIOR APPLICATION NUMBER: US 60/031,930
; PRIOR FILING DATE: 1996-11-27
; PRIOR APPLICATION NUMBER: US 60/048,547
; PRIOR FILING DATE: 1997-06-03
; NUMBER OF SEQ ID NOS: 53
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 44
; LENGTH: 546
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-671-242-44

```

```

Query Match 80.9%; Score 429; DB 12; Length 546;
Best Local Similarity 99.8%; Pred. No. 0;
Matches 529; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 KSSPGQPEAGPEGAQERPSQAAPAVEAEGPGSSQAAPRKEPGAQARTAQSGALRDVSEELS 60
DB 17 KSSPGQPEAGPEGAQERPSQAAPAVEAEGPGSSQAAPRKEPGAQARTAQSGALRDVSEELS 76
QY 61 ROLEDILSTYCVDNNGGPGEDGAQGEPAEPEDAESKSTYVARNGEPEPTPVVYGEKPS 120
DB 77 ROLEDILSTYCVDNNGGPGEDGAQGEPAEPEDAESKSTYVARNGEPEPTPVVYGEKPS 136
QY 121 KGPNTSEIRQSDVEGDRHRRPQEKKAAGLKEITLLMOTLNTLSTPEEKLAALCKKY 180
DB 137 KGPNTSEIRQSDVEGDRHRRPQEKKAAGLKEITLLMOTLNTLSTPEEKLAALCKKY 196
QY 181 AELLSEHNSQKQMLKQKQSQOLVQKHRLRGEHSHKAVLARSLKSLCRLQHRNRSIK 240
DB 197 AELLSEHNSQKQMLKQKQSQOLVQKHRLRGEHSHKAVLARSLKSLCRLQHRNRSIK 256
QY 241 EGVQARAESEKKEVTSHFQVTLNDIQLQMEQHNRNSKLRQENNELAERLKLIEQY 300
DB 257 EGVQARAESEKKEVTSHFQVTLNDIQLQMEQHNRNSKLRQENNELAERLKLIEQY 316

```

```

QY 301 ELREHIDKVFHKHLOQQOLVDKLOQAQEMLKEABERHOREKDFLLKEAVESQRMCELM 360
DB 317 ELREHIDKVFHKHLOQQOLVDKLOQAQEMLKEABERHOREKDFLLKEAVESQRMCELM 376
QY 361 KQETHLKQOLALYTEKPEEFQNTLSKSSSEVFTTFKQEMEKMTKKIKLEKETMYRSRW 420
DB 377 KQETHLKQOLALYTEKPEEFQNTLSKSSSEVFTTFKQEMEKMTKKIKLEKETMYRSRW 436
QY 421 ESSNKALLEMABEKTVRDKEGLEQVKIORLEKLCALQTERNDLNKRVODLSAGQGSLS 480
DB 437 ESSNKALLEMABEKTVRDKEGLEQVKIORLEKLCALQTERNDLNKRVODLSAGQGSLS 496
QY 481 TDSGPERRPEGGAQAPSSPRVTEAPCYPGAPSTEASGOTGPQEPSARA 530
DB 497 TDSGPERRPEGGAQAPSSPRVTEAPCYPGAPSTEASGOTGPQEPSARA 546

```

```

RESULT 9
US-10-023-529-44
; Sequence 44, Application US/10023529
; Publication No. US20020129388A1
; GENERAL INFORMATION:
; APPLICANT: Lees, Ann M.
; APPLICANT: Lees, Robert S.
; APPLICANT: Law, Simon W.
; APPLICANT: Arjona, Anibal A.
; TITLE OF INVENTION: NOVEL LOW DENSITY LIPOPROTEIN BINDING
; TITLE OF INVENTION: PROTEINS AND THEIR USE IN DIAGNOSING AND TREATING
; TITLE OF INVENTION: ATHEROSCLEROSIS
; FILE REFERENCE: 10797-004001
; CURRENT APPLICATION NUMBER: US/10/023,529
; CURRENT FILING DATE: 2001-12-17
; PRIOR APPLICATION NUMBER: 09/616,289
; PRIOR FILING DATE: 2000-07-14
; PRIOR APPLICATION NUMBER: US 09/517,849
; PRIOR FILING DATE: 2000-03-02
; PRIOR APPLICATION NUMBER: US 08/979,608
; PRIOR FILING DATE: 1997-11-26
; PRIOR APPLICATION NUMBER: US 60/031,930
; PRIOR FILING DATE: 1996-11-27
; PRIOR APPLICATION NUMBER: US 60/048,547
; PRIOR FILING DATE: 1997-06-03
; NUMBER OF SEQ ID NOS: 53
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 44
; LENGTH: 546
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-023-529-44

```

```

Query Match 80.9%; Score 429; DB 13; Length 546;
Best Local Similarity 99.8%; Pred. No. 0;
Matches 529; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 KSSPGQPEAGPEGAQERPSQAAPAVEAEGPGSSQAAPRKEPGAQARTAQSGALRDVSEELS 60
DB 17 KSSPGQPEAGPEGAQERPSQAAPAVEAEGPGSSQAAPRKEPGAQARTAQSGALRDVSEELS 76
QY 61 ROLEDILSTYCVDNNGGPGEDGAQGEPAEPEDAESKSTYVARNGEPEPTPVVYGEKPS 120
DB 77 ROLEDILSTYCVDNNGGPGEDGAQGEPAEPEDAESKSTYVARNGEPEPTPVVYGEKPS 136
QY 121 KGPNTSEIRQSDVEGDRHRRPQEKKAAGLKEITLLMOTLNTLSTPEEKLAALCKKY 180
DB 137 KGPNTSEIRQSDVEGDRHRRPQEKKAAGLKEITLLMOTLNTLSTPEEKLAALCKKY 196
QY 181 AELLSEHNSQKQMLKQKQSQOLVQKHRLRGEHSHKAVLARSLKSLCRLQHRNRSIK 240
DB 197 AELLSEHNSQKQMLKQKQSQOLVQKHRLRGEHSHKAVLARSLKSLCRLQHRNRSIK 256
QY 241 EGVQARAESEKKEVTSHFQVTLNDIQLQMEQHNRNSKLRQENNELAERLKLIEQY 300

```

Db 257 EGVQARAEERKEKVTSHFQVTLNDIQLQMEQHNERNSKLQENMELAEKLLIEQY 316
QY 301 ELREEHIDKVKFKHDLQOOLVDKLLQQAQEMLEKAEERHOREKDFLLKEAVESQRMCELM 360
Db 317 ELREEHIDKVKFKHDLQOOLVDKLLQQAQEMLEKAEERHOREKDFLLKEAVESQRMCELM 376
QY 361 KQOETHLKOQALALYTEKFEFQNTLSKSEVFYTFKQEMEMONTKKIKLEKETTYMYSRW 420
Db 377 KQOETHLKOQALALYTEKFEFQNTLSKSEVFYTFKQEMEMONTKKIKLEKETTYMYSRW 436
QY 421 ESSNKALLEMAEKTVDKLEGLQVKIORLEKLCRALQTERNDLNKRVODLSAGGGSLL 480
Db 437 ESSNKALLEMAEKTVDKLEGLQVKIORLEKLCRALQTERNDLNKRVODLSAGGGSLL 496
QY 481 TDSGPERRPEGGAQAPSSPRVTEAPCYPGAPSTEASQGTGPQPTTSARA 530
Db 497 TDSGPERRPEGGAQAPSSPRVTEAPCYPGAPSTEASQGTGPQPTTSARA 546

RESULT 10
US-10-023-523-44
; Sequence 44, Application US/10023523
; Publication No. US20020152485A1
; GENERAL INFORMATION:
; APPLICANT: Lees, Ann M.
; APPLICANT: Lees, Robert S.
; APPLICANT: Law, Simon W.
; APPLICANT: Arjona, Anibal A.
; TITLE OF INVENTION: NOVEL LOW DENSITY LIPOPROTEIN BINDING
; TITLE OF INVENTION: PROTEINS AND THEIR USE IN DIAGNOSING AND TREATING
; TITLE OF INVENTION: ATHEROSCLEROSIS
; FILE REFERENCE: 10797-004001
; CURRENT APPLICATION NUMBER: US/10/023,523
; CURRENT FILING DATE: 2001-12-17
; PRIOR APPLICATION NUMBER: US/09/616,289
; PRIOR FILING DATE: 2000-07-14
; PRIOR APPLICATION NUMBER: US 09/517,849
; PRIOR FILING DATE: 2000-03-02
; PRIOR APPLICATION NUMBER: US 08/979,608
; PRIOR FILING DATE: 1997-11-26
; PRIOR APPLICATION NUMBER: US 60/031,930
; PRIOR FILING DATE: 1996-11-27
; PRIOR APPLICATION NUMBER: US 60/048,547
; PRIOR FILING DATE: 1997-06-03
; NUMBER OF SEQ ID NOS: 53
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 44
; LENGTH: 546
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-023-523-44

Query Match 80.9%; Score 429; DB 13; Length 546;
Best Local Similarity 99.8%; Pred. No. 0;
Matches 529; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 KSPGQPEAGPEGAQERPSQAAPAVEAEGPGSSQAPRKPGEQAQARTAGSGLRDVSELS 60
Db 17 KSPGQPEAGPEGAQERPSQAAPAVEAEGPGSSQAPRKPGEQAQARTAGSGLRDVSELS 76
QY 61 RQLEDILSTVCVDDNQGCGEDGAGCEPAEPDAEKSRITYVARNGEPEPTPVVYGEKPS 120
Db 77 RQLEDILSTVCVDDNQGCGEDGAGCEPAEPDAEKSRITYVARNGEPEPTPVVYGEKPS 136
QY 121 KGPNTPEIRQSDVEGDRDHRHPQEKKAAGLGEITLLMQTLNTLSTPEKLAALCKY 180
Db 137 KGPNTPEIRQSDVEGDRDHRHPQEKKAAGLGEITLLMQTLNTLSTPEKLAALCKY 196
QY 181 AELLEHRNSOKMKLLQKQSQLVQEKDHLRGHSHKAVLARSKLESICRELQHRNSLK 240
Db 197 AELLEHRNSOKMKLLQKQSQLVQEKDHLRGHSHKAVLARSKLESICRELQHRNSLK 256
QY 241 EGVQARAEERKEKVTSHFQVTLNDIQLQMEQHNERNSKLQENMELAEKLLIEQY 300

Db 257 EGVQARAEERKEKVTSHFQVTLNDIQLQMEQHNERNSKLQENMELAEKLLIEQY 316
QY 301 ELREEHIDKVKFKHDLQOOLVDKLLQQAQEMLEKAEERHOREKDFLLKEAVESQRMCELM 360
Db 317 ELREEHIDKVKFKHDLQOOLVDKLLQQAQEMLEKAEERHOREKDFLLKEAVESQRMCELM 376
QY 361 KQOETHLKOQALALYTEKFEFQNTLSKSEVFYTFKQEMEMONTKKIKLEKETTYMYSRW 420
Db 377 KQOETHLKOQALALYTEKFEFQNTLSKSEVFYTFKQEMEMONTKKIKLEKETTYMYSRW 436
QY 421 ESSNKALLEMAEKTVDKLEGLQVKIORLEKLCRALQTERNDLNKRVODLSAGGGSLL 480
Db 437 ESSNKALLEMAEKTVDKLEGLQVKIORLEKLCRALQTERNDLNKRVODLSAGGGSLL 496
QY 481 TDSGPERRPEGGAQAPSSPRVTEAPCYPGAPSTEASQGTGPQPTTSARA 530
Db 497 TDSGPERRPEGGAQAPSSPRVTEAPCYPGAPSTEASQGTGPQPTTSARA 546

RESULT 11
US-10-616-187-44
; Sequence 44, Application US/10616187
; Publication No. US20040013668A1
; GENERAL INFORMATION:
; APPLICANT: Lees, Ann M.
; APPLICANT: Lees, Robert S.
; APPLICANT: Law, Simon W.
; APPLICANT: Arjona, Anibal A.
; TITLE OF INVENTION: NOVEL LOW DENSITY LIPOPROTEIN BINDING
; TITLE OF INVENTION: PROTEINS AND THEIR USE IN DIAGNOSING AND TREATING
; TITLE OF INVENTION: ATHEROSCLEROSIS
; FILE REFERENCE: 10797-004001
; CURRENT APPLICATION NUMBER: US/10/616,187
; CURRENT FILING DATE: 2003-07-09
; PRIOR APPLICATION NUMBER: US/09/616,289
; PRIOR FILING DATE: 2000-07-14
; PRIOR APPLICATION NUMBER: US 09/517,849
; PRIOR FILING DATE: 2000-03-02
; PRIOR APPLICATION NUMBER: US 08/979,608
; PRIOR FILING DATE: 1997-11-26
; PRIOR APPLICATION NUMBER: US 60/031,930
; PRIOR FILING DATE: 1996-11-27
; PRIOR APPLICATION NUMBER: US 60/048,547
; PRIOR FILING DATE: 1997-06-03
; NUMBER OF SEQ ID NOS: 53
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 44
; LENGTH: 546
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-616-187-44

Query Match 80.9%; Score 429; DB 15; Length 546;
Best Local Similarity 99.8%; Pred. No. 0;
Matches 529; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 KSPGQPEAGPEGAQERPSQAAPAVEAEGPGSSQAPRKPGEQAQARTAGSGLRDVSELS 60
Db 17 KSPGQPEAGPEGAQERPSQAAPAVEAEGPGSSQAPRKPGEQAQARTAGSGLRDVSELS 76
QY 61 RQLEDILSTVCVDDNQGCGEDGAGCEPAEPDAEKSRITYVARNGEPEPTPVVYGEKPS 120
Db 77 RQLEDILSTVCVDDNQGCGEDGAGCEPAEPDAEKSRITYVARNGEPEPTPVVYGEKPS 136
QY 121 KGPNTPEIRQSDVEGDRDHRHPQEKKAAGLGEITLLMQTLNTLSTPEKLAALCKY 180
Db 137 KGPNTPEIRQSDVEGDRDHRHPQEKKAAGLGEITLLMQTLNTLSTPEKLAALCKY 196
QY 181 AELLEHRNSOKMKLLQKQSQLVQEKDHLRGHSHKAVLARSKLESICRELQHRNSLK 240
Db 197 AELLEHRNSOKMKLLQKQSQLVQEKDHLRGHSHKAVLARSKLESICRELQHRNSLK 256

241 EGVORAREEKEKEVTSHPVNTLNDIQLQMEQHNRNSKLRQENNELAERLKLIEQY 300
Db EGVORAREEKEKEVTSHPVNTLNDIQLQMEQHNRNSKLRQENNELAERLKLIEQY 316
301 ELREEHIDKVFHKHDLQQLVDAKLQQAQEMLKEAERHOREKDFLLKEAVESQRMCELM 360
Db ELREEHIDKVFHKHDLQQLVDAKLQQAQEMLKEAERHOREKDFLLKEAVESQRMCELM 376
361 KOETHLKOQALALYTEKFPFQNTLSKSSEVFTTFKQEMKMTKKIKLEKETTYRSRW 420
Db KOETHLKOQALALYTEKFPFQNTLSKSSEVFTTFKQEMKMTKKIKLEKETTYRSRW 436
377 KOETHLKOQALALYTEKFPFQNTLSKSSEVFTTFKQEMKMTKKIKLEKETTYRSRW 440
Db KOETHLKOQALALYTEKFPFQNTLSKSSEVFTTFKQEMKMTKKIKLEKETTYRSRW 456
421 ESSNKALLEAEBEKTVDKELEGLQVVKIQRLEKLCRALQTERNDLNKRVQDLSAGGQSL 480
Db ESSNKALLEAEBEKTVDKELEGLQVVKIQRLEKLCRALQTERNDLNKRVQDLSAGGQSL 496
437 ESSNKALLEAEBEKTVDKELEGLQVVKIQRLEKLCRALQTERNDLNKRVQDLSAGGQSL 496
481 TDSGPRRPGGAGQAPSSPRVTAPCPYGPAPSTASGOTGPQBPPTSARA 530
Db TDSGPRRPGGAGQAPSSPRVTAPCPYGPAPSTASGOTGPQBPPTSARA 546
497 TDSGPRRPGGAGQAPSSPRVTAPCPYGPAPSTASGOTGPQBPPTSARA 546
RESULT 12
US-10-276-774-2134
; Sequence 2134, Application US/10276774
; Publication No. US20040053245A1
; GENERAL INFORMATION:
; APPLICANT: Hyseq, Inc.
; APPLICANT: Tang, Y. Tom et al
; TITLE OF INVENTION: No. US20040053245A1el Nucleic Acids and Polypeptides
; FILE REFERENCE: 21272-030
; CURRENT APPLICATION NUMBER: US/10/276,774
; CURRENT FILING DATE: 2002-11-18
; PRIOR APPLICATION NUMBER: 09/560,875
; PRIOR FILING DATE: 2000-04-27
; PRIOR APPLICATION NUMBER: 09/496,914
; PRIOR FILING DATE: 2000-02-03
; NUMBER OF SEQ ID NOS: 2700
; SOFTWARE: Custom
; SEQ ID NO 2134
; LENGTH: 510
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)..(510)
; OTHER INFORMATION: Xaa = any amino acid or nothing
US-10-276-774-2134
Query Match 67.7%; Score 359; DB 12; Length 510;
Best Local Similarity 99.8%; Pred. No. 0;
Matches 459; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 1 KSSPGQPEAGPEGAQERFSQAAPAVEAEGPGSSQAAPRKEGAQAARTAGSALRDVSELS 60
Db 28 KSSPGQPEAGPEGAQERFSQAAPAVEAEGPGSSQAAPRKEGAQAARTAGSALRDVSELS 87
QY 61 RQLEDILSTYCVDNNGQPGEDGAGCEBAEPEDAESKSTYVARNGEPPTPVNGEKEPS 120
Db 88 RQLEDILSTYCVDNNGQPGEDGAGCEBAEPEDAESKSTYVARNGEPPTPVNGEKEPS 147
QY 121 KGDPNTEIRQSDVEGDRHRRPQKXKAGLGKEITLLMOTLNTLSTPEEKLALCKKY 180
Db 148 KGDPNTEIRQSDVEGDRHRRPQKXKAGLGKEITLLMOTLNTLSTPEEKLALCKKY 207
QY 181 AELLBEHRNSQOMKLLQKQSQLVQEKDHLGESHKAVLARSLSLRELOHNRSLK 240
Db 208 AELLBEHRNSQOMKLLQKQSQLVQEKDHLGESHKAVLARSLSLRELOHNRSLK 267
QY 241 EGVORAREEKEKEVTSHPVNTLNDIQLQMEQHNRNSKLRQENNELAERLKLIEQY 300
Db 268 EGVORAREEKEKEVTSHPVNTLNDIQLQMEQHNRNSKLRQENNELAERLKLIEQY 327
QY 301 ELREEHIDKVFHKHDLQQLVDAKLQQAQEMLKEAERHOREKDFLLKEAVESQRMCELM 360

328 ELREEHIDKVFHKHDLQQLVDAKLQQAQEMLKEAERHOREKDFLLKEAVESQRMCELM 387
QY 361 KOETHLKOQALALYTEKFPFQNTLSKSSEVFTTFKQEMKMTKKIKLEKETTYRSRW 420
Db 388 KOETHLKOQALALYTEKFPFQNTLSKSSEVFTTFKQEMKMTKKIKLEKETTYRSRW 447
QY 421 ESSNKALLEAEBEKTVDKELEGLQVVKIQRLEKLCRALQTERNDLNKRVQDLSAGGQSL 480
Db 448 ESSNKALLEAEBEKTVDKELEGLQVVKIQRLEKLCRALQTERNDLNKRVQDLSAGGQSL 496
RESULT 13
US-09-962-055-5
; Sequence 5, Application US/09962055
; Patent No. US20020052033A1
; GENERAL INFORMATION:
; APPLICANT: Lees, Ann M. S.
; APPLICANT: Lees, Robert S.
; APPLICANT: Law, Simon W.
; APPLICANT: Arjona, Anibal A.
; TITLE OF INVENTION: NOVEL LOW DENSITY LIPOPROTEIN
; BINDING PROTEINS AND THEIR USES IN DIAGNOSING AND
; TREATING ATHEROSCLEROSIS
; NUMBER OF SEQUENCES: 42
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Fish & Richardson P.C.
; STREET: 225 Franklin Street
; CITY: Boston
; STATE: MA
; COUNTRY: USA
; ZIP: 02110-2804
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/962,055
; FILING DATE: 24-Sep-2001
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/979,608
; FILING DATE: 26-NOV-1997
; APPLICATION NUMBER: US 60/031,930
; FILING DATE: 27-NOV-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: Myers, Louis
; REGISTRATION NUMBER: 35,965
; REFERENCE/DOCKET NUMBER: 10797-002001 (formerly 3983/59818)
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 617/542-5070
; TELEFAX: 617/542-8906
; INFORMATION FOR SEQ ID NO: 5:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 557 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; FRAGMENT TYPE: internal
; SEQUENCE DESCRIPTION: SEQ ID NO: 5:
US-09-962-055-5
Query Match 32.5%; Score 172; DB 9; Length 557;
Best Local Similarity 100.0%; Pred. No. 3,1e-155;
Matches 172; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 264 TLNDIQLQMEQHNRNSKLRQENNELAERLKLIEQYELREEHIDKVFHKHDLQQLVDA 323
Db 280 TLNDIQLQMEQHNRNSKLRQENNELAERLKLIEQYELREEHIDKVFHKHDLQQLVDA 339
QY 324 KLOQAQEMLKEAERHOREKDFLLKEAVESQRMCELMKQETHLKOQALALYTEKFPFQ 383
Db 340 KLOQAQEMLKEAERHOREKDFLLKEAVESQRMCELMKQETHLKOQALALYTEKFPFQ 399

```
QY 384 T L S K S S E V T T F K Q E M E K M T K I K L E K E T T M Y R S W S S N K A L L E M A E E K T 435
DB 400 T L S K S S E V T T F K Q E M E K M T K I K L E K E T T M Y R S W S S N K A L L E M A E E K T 451

RESULT 14
US-09-976-740-5
; Sequence 5, Application US/09976740
; Publication No. US20020194633A1
; GENERAL INFORMATION:
; APPLICANT: Lees, Robert S.
; APPLICANT: Law, Simon W.
; APPLICANT: Arjona, Anibal A.
; TITLE OF INVENTION: NOVEL LOW DENSITY LIPOPROTEIN BINDING
; TITLE OF INVENTION: PROTEINS AND THEIR USE IN DIAGNOSING AND TREATING
; TITLE OF INVENTION: ATHEROSCLEROSIS
; FILE REFERENCE: 10797-004001
; CURRENT APPLICATION NUMBER: US/09/976,740
; CURRENT FILING DATE: 2001-10-12
; PRIOR APPLICATION NUMBER: 09/616,289
; PRIOR FILING DATE: 2000-07-14
; PRIOR APPLICATION NUMBER: US 08/979,608
; PRIOR FILING DATE: 1997-11-26
; PRIOR APPLICATION NUMBER: US 60/031,930
; PRIOR FILING DATE: 1996-11-27
; PRIOR APPLICATION NUMBER: US 60/048,547
; PRIOR FILING DATE: 1997-06-03
; NUMBER OF SEQ ID NOS: 53
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 5
; LENGTH: 557
; TYPE: PRT
; ORGANISM: Oryctolagus cuniculus
US-09-976-740-5

Query Match 32.5%; Score 172; DB 9; Length 557;
Best Local Similarity 100.0%; Pred. No. 3.1e-155; Indels 0; Gaps 0;
Matches 172; Conservative 0; Mismatches 0;

QY 264 T L N D I Q L Q M E Q H N R N S K L R Q E N M E L A E R L K K L I E Q Y E L R E H I D K V F K H D L Q Q L V D A 323
DB 280 T L N D I Q L Q M E Q H N R N S K L R Q E N M E L A E R L K K L I E Q Y E L R E H I D K V F K H D L Q Q L V D A 339
QY 324 K L Q Q A Q E M L K E A E R H Q R E K D F L L K E A V E S Q R M C E L M K Q Q E T H L K Q Q L A L Y T E K P E E F Q N 383
DB 340 K L Q Q A Q E M L K E A E R H Q R E K D F L L K E A V E S Q R M C E L M K Q Q E T H L K Q Q L A L Y T E K P E E F Q N 399
QY 384 T L S K S S E V T T F K Q E M E K M T K I K L E K E T T M Y R S W S S N K A L L E M A E E K T 435
DB 400 T L S K S S E V T T F K Q E M E K M T K I K L E K E T T M Y R S W S S N K A L L E M A E E K T 451

RESULT 15
US-10-671-242-5
; Sequence 5, Application US/10671242
; Publication No. US20040040049A1
; GENERAL INFORMATION:
; APPLICANT: Lees, Robert S.
; APPLICANT: Law, Simon W.
; APPLICANT: Arjona, Anibal A.
; TITLE OF INVENTION: NOVEL LOW DENSITY LIPOPROTEIN BINDING
; TITLE OF INVENTION: PROTEINS AND THEIR USE IN DIAGNOSING AND TREATING
; TITLE OF INVENTION: ATHEROSCLEROSIS
; FILE REFERENCE: 10797-004001
; CURRENT APPLICATION NUMBER: US/10/671,242
; CURRENT FILING DATE: 2003-09-24
; PRIOR APPLICATION NUMBER: US/09/616,289
; PRIOR FILING DATE: 2000-07-14
; PRIOR APPLICATION NUMBER: US 09/517,849
; PRIOR FILING DATE: 2000-03-02
```

```
; PRIOR APPLICATION NUMBER: US 08/979,608
; PRIOR FILING DATE: 1997-11-26
; PRIOR APPLICATION NUMBER: US 60/031,930
; PRIOR FILING DATE: 1996-11-27
; PRIOR APPLICATION NUMBER: US 60/048,547
; PRIOR FILING DATE: 1997-06-03
; NUMBER OF SEQ ID NOS: 53
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 5
; LENGTH: 557
; TYPE: PRT
; ORGANISM: Oryctolagus cuniculus
US-10-671-242-5

Query Match 32.5%; Score 172; DB 12; Length 557;
Best Local Similarity 100.0%; Pred. No. 3.1e-155; Indels 0; Gaps 0;
Matches 172; Conservative 0; Mismatches 0;

QY 264 T L N D I Q L Q M E Q H N R N S K L R Q E N M E L A E R L K K L I E Q Y E L R E H I D K V F K H D L Q Q L V D A 323
DB 280 T L N D I Q L Q M E Q H N R N S K L R Q E N M E L A E R L K K L I E Q Y E L R E H I D K V F K H D L Q Q L V D A 339
QY 324 K L Q Q A Q E M L K E A E R H Q R E K D F L L K E A V E S Q R M C E L M K Q Q E T H L K Q Q L A L Y T E K P E E F Q N 383
DB 340 K L Q Q A Q E M L K E A E R H Q R E K D F L L K E A V E S Q R M C E L M K Q Q E T H L K Q Q L A L Y T E K P E E F Q N 399
QY 384 T L S K S S E V T T F K Q E M E K M T K I K L E K E T T M Y R S W S S N K A L L E M A E E K T 435
DB 400 T L S K S S E V T T F K Q E M E K M T K I K L E K E T T M Y R S W S S N K A L L E M A E E K T 451

RESULT 16
US-10-023-529-5
; Sequence 5, Application US/10023529
; Publication No. US20020129388A1
; GENERAL INFORMATION:
; APPLICANT: Lees, Robert S.
; APPLICANT: Law, Simon W.
; APPLICANT: Arjona, Anibal A.
; TITLE OF INVENTION: NOVEL LOW DENSITY LIPOPROTEIN BINDING
; TITLE OF INVENTION: PROTEINS AND THEIR USE IN DIAGNOSING AND TREATING
; TITLE OF INVENTION: ATHEROSCLEROSIS
; FILE REFERENCE: 10797-004001
; CURRENT APPLICATION NUMBER: US/10/023,529
; CURRENT FILING DATE: 2001-12-17
; PRIOR APPLICATION NUMBER: 09/616,289
; PRIOR FILING DATE: 2000-07-14
; PRIOR APPLICATION NUMBER: US 09/517,849
; PRIOR FILING DATE: 2000-03-02
; PRIOR APPLICATION NUMBER: US 08/979,608
; PRIOR FILING DATE: 1997-11-26
; PRIOR APPLICATION NUMBER: US 60/031,930
; PRIOR FILING DATE: 1996-11-27
; PRIOR APPLICATION NUMBER: US 60/048,547
; PRIOR FILING DATE: 1997-06-03
; NUMBER OF SEQ ID NOS: 53
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 5
; LENGTH: 557
; TYPE: PRT
; ORGANISM: Oryctolagus cuniculus
US-10-023-529-5

Query Match 32.5%; Score 172; DB 13; Length 557;
Best Local Similarity 100.0%; Pred. No. 3.1e-155; Indels 0; Gaps 0;
Matches 172; Conservative 0; Mismatches 0;

QY 264 T L N D I Q L Q M E Q H N R N S K L R Q E N M E L A E R L K K L I E Q Y E L R E H I D K V F K H D L Q Q L V D A 323
DB 280 T L N D I Q L Q M E Q H N R N S K L R Q E N M E L A E R L K K L I E Q Y E L R E H I D K V F K H D L Q Q L V D A 339
QY 324 K L Q Q A Q E M L K E A E R H Q R E K D F L L K E A V E S Q R M C E L M K Q Q E T H L K Q Q L A L Y T E K P E E F Q N 383
```

```
Db 340 KLOQAQEMLEKAEERHOREKDFLLKEAVESQRCMELMKQETHLKQQLALYTEKPEEPQN 399
OY 384 TLSKSEVFTTFKQEMKMTKKIKLEKETTYRSRWSNKKALLEMAEKT 435
Db 400 TLSKSEVFTTFKQEMKMTKKIKLEKETTYRSRWSNKKALLEMAEKT 451

RESULT 17
US-10-023-523-5
; Sequence 5, Application US/10023523
; Publication No. US20020152485A1
; GENERAL INFORMATION:
; APPLICANT: Lees, Ann M.
; APPLICANT: Lees, Robert S.
; APPLICANT: Law, Simon W.
; APPLICANT: Arjona, Anibal A.
; TITLE OF INVENTION: NOVEL LOW DENSITY LIPOPROTEIN BINDING
; TITLE OF INVENTION: PROTEINS AND THEIR USE IN DIAGNOSING AND TREATING
; FILE REFERENCE: ATHEROSCLEROSIS
; CURRENT APPLICATION NUMBER: US/10/023,523
; PRIOR FILING DATE: 2001-12-17
; PRIOR APPLICATION NUMBER: US/09/616,289
; PRIOR FILING DATE: 2000-07-14
; PRIOR APPLICATION NUMBER: US 09/517,849
; PRIOR FILING DATE: 2000-03-02
; PRIOR APPLICATION NUMBER: US 08/979,608
; PRIOR FILING DATE: 1997-11-26
; PRIOR APPLICATION NUMBER: US 60/031,930
; PRIOR FILING DATE: 1996-11-27
; PRIOR APPLICATION NUMBER: US 60/048,547
; PRIOR FILING DATE: 1997-06-03
; NUMBER OF SEQ ID NOS: 53
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 5
; LENGTH: 557
; TYPE: PRT
; ORGANISM: Oryctolagus cuniculus
US-10-023-523-5

Query Match 32.5%; Score 172; DB 13; Length 557;
Best Local Similarity 100.0%; Pred. No. 3.1e-155;
Matches 172; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 264 TLNDIQLQWQHNRNSKLRQENMELAEKLLIQYELREHIDKVPFKHDLQQLVDA 323
Db 280 TLNDIQLQWQHNRNSKLRQENMELAEKLLIQYELREHIDKVPFKHDLQQLVDA 339
OY 324 KLOQAQEMLEKAEERHOREKDFLLKEAVESQRCMELMKQETHLKQQLALYTEKPEEPQN 383
Db 340 KLOQAQEMLEKAEERHOREKDFLLKEAVESQRCMELMKQETHLKQQLALYTEKPEEPQN 399
OY 384 TLSKSEVFTTFKQEMKMTKKIKLEKETTYRSRWSNKKALLEMAEKT 435
Db 400 TLSKSEVFTTFKQEMKMTKKIKLEKETTYRSRWSNKKALLEMAEKT 451

RESULT 19
US-10-023-529-8
; Sequence 3413, Application US/10264049
; Publication No. US20040005579A1
; GENERAL INFORMATION:
; APPLICANT: Birse et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: PA133PI
; CURRENT APPLICATION NUMBER: US/10/264,049
; CURRENT FILING DATE: 2002-10-04
; PRIOR APPLICATION NUMBER: PCT/US01/18569
; PRIOR FILING DATE: 2001-06-07
; PRIOR APPLICATION NUMBER: US 60/209,467
; PRIOR FILING DATE: 2000-06-07
; NUMBER OF SEQ ID NOS: 4360
; SOFTWARE: PatentIn Ver. 3.1
; SEQ ID NO 3413
; LENGTH: 85
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: MISC_FEATURE
; LOCATION: (1)
; OTHER INFORMATION: Xaa equals any of the twenty naturally occurring L-amino acids
; FEATURE:
; NAME/KEY: MISC_FEATURE
; LOCATION: (3)
; OTHER INFORMATION: Xaa equals any of the twenty naturally occurring L-amino acids
; FEATURE:
; NAME/KEY: MISC_FEATURE
; LOCATION: (18)
; OTHER INFORMATION: Xaa equals any of the twenty naturally occurring L-amino acids
; FEATURE:
; NAME/KEY: MISC_FEATURE
; LOCATION: (75)
; OTHER INFORMATION: Xaa equals any of the twenty naturally occurring L-amino acids
US-10-023-529-8

Query Match 32.5%; Score 172; DB 13; Length 557;
Best Local Similarity 100.0%; Pred. No. 3.1e-155;
Matches 172; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 264 TLNDIQLQWQHNRNSKLRQENMELAEKLLIQYELREHIDKVPFKHDLQQLVDA 323
Db 280 TLNDIQLQWQHNRNSKLRQENMELAEKLLIQYELREHIDKVPFKHDLQQLVDA 339
OY 324 KLOQAQEMLEKAEERHOREKDFLLKEAVESQRCMELMKQETHLKQQLALYTEKPEEPQN 383
Db 340 KLOQAQEMLEKAEERHOREKDFLLKEAVESQRCMELMKQETHLKQQLALYTEKPEEPQN 399
OY 384 TLSKSEVFTTFKQEMKMTKKIKLEKETTYRSRWSNKKALLEMAEKT 435
Db 400 TLSKSEVFTTFKQEMKMTKKIKLEKETTYRSRWSNKKALLEMAEKT 451

RESULT 18
US-10-023-529-8
; Sequence 5, Application US/10616187
; Publication No. US20040013669A1
; GENERAL INFORMATION:
; APPLICANT: Lees, Ann M.
; APPLICANT: Lees, Robert S.
; APPLICANT: Law, Simon W.
; APPLICANT: Arjona, Anibal A.
; TITLE OF INVENTION: NOVEL LOW DENSITY LIPOPROTEIN BINDING
; TITLE OF INVENTION: PROTEINS AND THEIR USE IN DIAGNOSING AND TREATING
; FILE REFERENCE: ATHEROSCLEROSIS
; CURRENT APPLICATION NUMBER: US/10/616,187
; CURRENT FILING DATE: 2003-07-09
```

```
; PRIOR APPLICATION NUMBER: US/09/616,289
; PRIOR FILING DATE: 2000-07-14
; PRIOR APPLICATION NUMBER: US 09/517,849
; PRIOR FILING DATE: 2000-03-02
; PRIOR APPLICATION NUMBER: US 08/979,608
; PRIOR FILING DATE: 1997-11-26
; PRIOR APPLICATION NUMBER: US 60/031,930
; PRIOR FILING DATE: 1996-11-27
; PRIOR APPLICATION NUMBER: US 60/048,547
; PRIOR FILING DATE: 1997-06-03
; NUMBER OF SEQ ID NOS: 53
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 5
; LENGTH: 557
; TYPE: PRT
; ORGANISM: Oryctolagus cuniculus
US-10-023-523-5

Query Match 32.5%; Score 172; DB 15; Length 557;
Best Local Similarity 100.0%; Pred. No. 3.1e-155;
Matches 172; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 264 TLNDIQLQWQHNRNSKLRQENMELAEKLLIQYELREHIDKVPFKHDLQQLVDA 323
Db 280 TLNDIQLQWQHNRNSKLRQENMELAEKLLIQYELREHIDKVPFKHDLQQLVDA 339
OY 324 KLOQAQEMLEKAEERHOREKDFLLKEAVESQRCMELMKQETHLKQQLALYTEKPEEPQN 383
Db 340 KLOQAQEMLEKAEERHOREKDFLLKEAVESQRCMELMKQETHLKQQLALYTEKPEEPQN 399
OY 384 TLSKSEVFTTFKQEMKMTKKIKLEKETTYRSRWSNKKALLEMAEKT 435
Db 400 TLSKSEVFTTFKQEMKMTKKIKLEKETTYRSRWSNKKALLEMAEKT 451

RESULT 19
US-10-0264-049-3413
; Sequence 3413, Application US/10264049
; Publication No. US20040005579A1
; GENERAL INFORMATION:
; APPLICANT: Birse et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: PA133PI
; CURRENT APPLICATION NUMBER: US/10/264,049
; CURRENT FILING DATE: 2002-10-04
; PRIOR APPLICATION NUMBER: PCT/US01/18569
; PRIOR FILING DATE: 2001-06-07
; PRIOR APPLICATION NUMBER: US 60/209,467
; PRIOR FILING DATE: 2000-06-07
; NUMBER OF SEQ ID NOS: 4360
; SOFTWARE: PatentIn Ver. 3.1
; SEQ ID NO 3413
; LENGTH: 85
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: MISC_FEATURE
; LOCATION: (1)
; OTHER INFORMATION: Xaa equals any of the twenty naturally occurring L-amino acids
; FEATURE:
; NAME/KEY: MISC_FEATURE
; LOCATION: (3)
; OTHER INFORMATION: Xaa equals any of the twenty naturally occurring L-amino acids
; FEATURE:
; NAME/KEY: MISC_FEATURE
; LOCATION: (18)
; OTHER INFORMATION: Xaa equals any of the twenty naturally occurring L-amino acids
; FEATURE:
; NAME/KEY: MISC_FEATURE
; LOCATION: (75)
; OTHER INFORMATION: Xaa equals any of the twenty naturally occurring L-amino acids
US-10-264-049-3413
```

Query Match 3.2%; Score 17; DB 15; Length 85;
Best Local Similarity 100.0%; Pred. No. 6.3e-08;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 397 QEMEKTWKIKKLEKET 413
|||||
DB 52 QEMEKTWKIKKLEKET 68

RESULT 20
US-10-424-599-207434
; Sequence 207434, Application US/10424599
; Publication No. US20040031072A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa Thomas J
; APPLICANT: Kovalic David K
; APPLICANT: Zhou Yihua
; APPLICANT: Cao Yongwei
; TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated with
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
; FILE REFERENCE: 38-21(53223)B
; CURRENT APPLICATION NUMBER: US/10/424,599
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 285684
; SEQ ID NO 207434
; LENGTH: 436
; TYPE: PRT
; ORGANISM: Glycine max
; FEATURE:
; NAME/KEY: unsure
; LOCATION: (1)...(436)
; OTHER INFORMATION: unsure at all Xaa locations
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT3847_29338C.1.pep
US-10-424-599-207434

Query Match 2.1%; Score 11; DB 12; Length 436;
Best Local Similarity 100.0%; Pred. No. 0.15;
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 224 KLESCLRELQR 234
|||||
DB 194 KLESCLRELQR 204

RESULT 21
US-09-764-869-1082
; Sequence 1082, Application US/09764869
; Patent No. US20020061521A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: PC007
; CURRENT APPLICATION NUMBER: US/09/764,869
; CURRENT FILING DATE: 2001-01-17
; Prior application data removed - refer to PALM or file wrapper
; NUMBER OF SEQ ID NOS: 2442
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1082
; LENGTH: 99
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: SITE
; LOCATION: (45)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
; NAME/KEY: SITE
; LOCATION: (90)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
; NAME/KEY: SITE
; LOCATION: (97)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
US-09-764-869-1082

Query Match 1.7%; Score 9; DB 9; Length 99;
Best Local Similarity 100.0%; Pred. No. 3.2;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 115 GEKEPSKGD 123
|||||
DB 49 GEKEPSKGD 57

RESULT 22
US-10-091-504-1082
; Sequence 1082, Application US/10091504
; Publication No. US2003005908A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: PC007C1
; CURRENT APPLICATION NUMBER: US/10/091,504
; CURRENT FILING DATE: 2002-03-07
; NUMBER OF SEQ ID NOS: 2442
; Prior Application removed - See File Wrapper or Palm
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1082
; LENGTH: 99
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (45)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
; NAME/KEY: misc feature
; LOCATION: (90)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
; NAME/KEY: misc feature
; LOCATION: (97)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
US-10-091-504-1082

Query Match 1.7%; Score 9; DB 14; Length 99;
Best Local Similarity 100.0%; Pred. No. 3.2;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 115 GEKEPSKGD 123
|||||
DB 49 GEKEPSKGD 57

RESULT 23
US-10-227-577-1082
; Sequence 1082, Application US/10227577
; Publication No. US20040005575A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: PC007C2
; CURRENT APPLICATION NUMBER: US/10/227,577
; CURRENT FILING DATE: 2002-08-26
; Prior Application Number: 10/091,504
; Prior Filing Date: 2002-03-07
; Prior Application Number: 09/764,869
; Prior Filing Date: 2001-01-17
; Prior Application Number: 60/179,065
; Prior Filing Date: 2000-01-31
; Prior Application Number: 60/180,628
; Prior Filing Date: 2000-02-04
; Prior Application Number: 60/214,886
; Prior Filing Date: 2000-06-28
; Prior Application Number: 60/217,487
; Prior Filing Date: 2000-07-11
; Prior Application Number: 60/225,758
; Prior Filing Date: 2000-08-14
; Prior Application Number: 60/220,963

PRIOR FILING DATE: 2000-07-26
PRIOR APPLICATION NUMBER: 60/217,496
PRIOR FILING DATE: 2000-07-21
PRIOR APPLICATION NUMBER: 60/225,447
PRIOR FILING DATE: 2000-08-14
Remaining Prior Application data removed - See File Wrapper or PALM.
NUMBER OF SEQ ID NOS: 2442
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 1082
LENGTH: 99
TYPE: PRT
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: misc feature
LOCATION: (45)
OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
FEATURE:
NAME/KEY: misc feature
LOCATION: (90)
OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
FEATURE:
NAME/KEY: misc feature
LOCATION: (97)
OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
US-10-227-577-1082

Query Match 1.7%; Score 9; DB 15; Length 99;
Best Local Similarity 100.0%; Pred. No. 3.2;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 115 GEKEPSKGD 123
Db 49 GEKEPSKGD 57

RESULT 24
US-09-801-368-112
Sequence 112, Application US/09801368
Patent No. US20020128250A1
GENERAL INFORMATION:
APPLICANT: Busby, Robert
APPLICANT: Cali, Brian
APPLICANT: Hecht, Peter
APPLICANT: Holtzman, Doug
APPLICANT: Madden, Kevin
APPLICANT: Maxon, Mary
APPLICANT: Milne, Todd
APPLICANT: No. US20020128250A1man, Thea
APPLICANT: Royer, John
APPLICANT: Salama, Sofie
APPLICANT: Sherman, Amir
APPLICANT: Silva, Jeff
APPLICANT: Summers, Eric
TITLE OF INVENTION: Methods for Improving Secondary Metabolite Production in Fungi
FILE REFERENCE: 109272.147
CURRENT APPLICATION NUMBER: US/09/801.368
CURRENT FILING DATE: 2001-03-07
PRIOR APPLICATION NUMBER: US 09/487,558
PRIOR FILING DATE: 2000-01-19
PRIOR APPLICATION NUMBER: US 60/160,587
PRIOR FILING DATE: 1999-10-20
NUMBER OF SEQ ID NOS: 440
SOFTWARE: PatentIn version 3.0
SEQ ID NO 112
LENGTH: 594
TYPE: PRT
ORGANISM: Saccharomyces cerevisiae
US-09-801-368-112

Query Match 1.7%; Score 9; DB 9; Length 594;
Best Local Similarity 100.0%; Pred. No. 16;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 164 NTLSTPEEK 172
Db 466 NTLSTPEEK 474
RESULT 25
US-10-369-493-1762
Sequence 1762, Application US/10369493
Publication No. US20030233675A1
GENERAL INFORMATION:
APPLICANT: Cao, Youngwei
APPLICANT: Hinkle, Gregory J.
APPLICANT: Slater, Steven C.
APPLICANT: Goldman, Barry S.
APPLICANT: Chen, Xianfeng
TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF
FILE REFERENCE: 38-10(52052)B
CURRENT APPLICATION NUMBER: US/10/369,493
CURRENT FILING DATE: 2003-02-28
PRIOR APPLICATION NUMBER: US 60/360,039
PRIOR FILING DATE: 2002-02-21
NUMBER OF SEQ ID NOS: 47374
SEQ ID NO 1762
LENGTH: 594
TYPE: PRT
ORGANISM: Saccharomyces cerevisiae
US-10-369-493-1762

Query Match 1.7%; Score 9; DB 15; Length 594;
Best Local Similarity 100.0%; Pred. No. 16;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 164 NTLSTPEEK 172
Db 466 NTLSTPEEK 474

RESULT 26
US-10-282-122A-72115
Sequence 72115, Application US/10282122A
Publication No. US20040029129A1
GENERAL INFORMATION:
APPLICANT: Wang, Liangeu
APPLICANT: Zamudio, Carlos
APPLICANT: Malone, Cheryl
APPLICANT: Haselbeck, Robert
APPLICANT: Ohlsen, Kari
APPLICANT: Zyskind, Judith
APPLICANT: Wall, Daniel
APPLICANT: Trawick, John
APPLICANT: Carr, Grant
APPLICANT: Yamamoto, Robert
APPLICANT: Forsyth, R.
APPLICANT: Xu, H.
TITLE OF INVENTION: Identification of Essential Genes in Microorganisms
FILE REFERENCE: ELITRA.034A
CURRENT APPLICATION NUMBER: US/10/282,122A
CURRENT FILING DATE: 2003-02-20
PRIOR APPLICATION NUMBER: 60/191,078
PRIOR FILING DATE: 2000-03-21
PRIOR APPLICATION NUMBER: 60/206,848
PRIOR FILING DATE: 2000-05-23
PRIOR APPLICATION NUMBER: 60/207,727
PRIOR FILING DATE: 2000-05-26
PRIOR APPLICATION NUMBER: 60/230,335
PRIOR FILING DATE: 2000-09-06
PRIOR APPLICATION NUMBER: 60/230,347
PRIOR FILING DATE: 2000-09-09
PRIOR APPLICATION NUMBER: 60/242,578
PRIOR FILING DATE: 2000-10-23
PRIOR APPLICATION NUMBER: 60/253,625
PRIOR FILING DATE: 2000-11-27

; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/267,636
; PRIOR FILING DATE: 2001-02-09
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 78614
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 72115
; LENGTH: 637
; TYPE: PRT
; ORGANISM: Streptococcus mutans
US-10-282-122A-72115

Query Match 1.7%; Score 9; DB 12; Length 637;
Best Local Similarity 100.0%; Pred. No. 17;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 57 BELSRQLED 65
Db 629 BELSRQLED 637

RESULT 27
US-10-443-622-130
; Sequence 130, Application US/10443622
; Publication No. US20040024192A1
; GENERAL INFORMATION:
; APPLICANT: Carter et al.
; TITLE OF INVENTION: 19 Human Secreted Proteins
; FILE REFERENCE: P2009P1
; CURRENT APPLICATION NUMBER: US/10/443,622
; CURRENT FILING DATE: 2003-05-23
; PRIOR APPLICATION NUMBER: PCT/US98/13608
; PRIOR FILING DATE: 1998-06-30
; PRIOR APPLICATION NUMBER: 60/051,480
; PRIOR FILING DATE: 1997-07-01
; PRIOR APPLICATION NUMBER: 60/051,381
; PRIOR FILING DATE: 1997-07-01
; PRIOR APPLICATION NUMBER: 60/058,663
; PRIOR FILING DATE: 1997-09-12
; PRIOR APPLICATION NUMBER: 60/058,598
; PRIOR FILING DATE: 1997-09-12
; NUMBER OF SEQ ID NOS: 156
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 130
; LENGTH: 25
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-443-622-130

Query Match 1.5%; Score 8; DB 16; Length 25;
Best Local Similarity 100.0%; Pred. No. 8.4;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 468 RVQDLSAG 475
Db 5 RVQDLSAG 12

RESULT 28
US-10-425-114-4846
; Sequence 4846, Application US/10425114
; Publication No. US20040034888A1
; GENERAL INFORMATION:
; APPLICANT: Liu, Jingdong
; APPLICANT: Zhou, Yihua
; APPLICANT: Kovalic, David K.
; APPLICANT: Screen, Steven E.
; APPLICANT: Tabaska, Jack E.
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With

; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
; FILE REFERENCE: 38-21(5313)B
; CURRENT APPLICATION NUMBER: US/10/425,114
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 73128
; SEQ ID NO 48446
; LENGTH: 106
; TYPE: PRT
; ORGANISM: Zea mays
; FEATURE:
; OTHER INFORMATION: Clone ID: 700041743_FLI.pep
US-10-425-114-4846

Query Match 1.5%; Score 8; DB 12; Length 106;
Best Local Similarity 100.0%; Pred. No. 30;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 10 GPEGAQER 17
Db 27 GPEGAQER 34

RESULT 29
US-10-424-599-272074
; Sequence 272074, Application US/10424599
; Publication No. US20040031072A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa Thomas J
; APPLICANT: Kovalic David K
; APPLICANT: Zhou Yihua
; APPLICANT: Cao Yongwei
; TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated With
; FILE REFERENCE: 38-21(53223)B
; CURRENT APPLICATION NUMBER: US/10/424,599
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 285684
; SEQ ID NO 272074
; LENGTH: 121
; TYPE: PRT
; ORGANISM: Glycine max
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT3847_87701C.1.pep
US-10-424-599-272074

Query Match 1.5%; Score 8; DB 12; Length 121;
Best Local Similarity 100.0%; Pred. No. 34;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 30 PGSSQAPR 37
Db 70 PGSSQAPR 77

RESULT 30
US-10-443-622-122
; Sequence 122, Application US/10443622
; Publication No. US20040024192A1
; GENERAL INFORMATION:
; APPLICANT: Carter et al.
; TITLE OF INVENTION: 19 Human Secreted Proteins
; FILE REFERENCE: P2009P1
; CURRENT APPLICATION NUMBER: US/10/443,622
; CURRENT FILING DATE: 2003-05-23
; PRIOR APPLICATION NUMBER: PCT/US98/13608
; PRIOR FILING DATE: 1998-06-30
; PRIOR APPLICATION NUMBER: 60/051,480
; PRIOR FILING DATE: 1997-07-01
; PRIOR APPLICATION NUMBER: 60/051,381
; PRIOR FILING DATE: 1997-07-01
; PRIOR APPLICATION NUMBER: 60/058,663
; PRIOR FILING DATE: 1997-09-12
; PRIOR APPLICATION NUMBER: 60/058,598


```
; PRIOR FILING DATE: 1997-09-12
; NUMBER OF SEQ ID NOS: 156
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 122
; LENGTH: 156
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-443-622-122

Query Match      1.5%; Score 8; DB 16; Length 156;
Best Local Similarity 100.0%; Pred. No. 43;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy      468 RVQDLSAG 475
Db      87 RVQDLSAG 94

RESULT 31
US-10-080-170-316
; Sequence 316, Application US/10080170
; Publication No. US20030129601A1
; GENERAL INFORMATION:
; APPLICANT: COLE, S.T.
; TITLE OF INVENTION: COMPARATIVE MYCOBACTERIAL GENOMICS AS A TOOL FOR
; IDENTIFYING TARGETS FOR THE DIAGNOSIS, PROPHYLAXIS OR
; TREATMENT OF MYCOBACTERIOSES
; TITLE OF INVENTION: TREATMENT OF MYCOBACTERIOSES
; FILE REFERENCE: 03495-0218
; CURRENT APPLICATION NUMBER: US/10/080,170
; CURRENT FILING DATE: 2002-06-10
; PRIOR APPLICATION NUMBER: 60/270,123
; PRIOR FILING DATE: 2001-02-22
; NUMBER OF SEQ ID NOS: 652
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 316
; LENGTH: 229
; TYPE: PRT
; ORGANISM: Mycobacterium leprae
US-10-080-170-316

Query Match      1.5%; Score 8; DB 14; Length 229;
Best Local Similarity 100.0%; Pred. No. 60;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy      18 PSQAAPAV 25
Db      200 PSQAAPAV 207

RESULT 32
US-10-156-761-12433
; Sequence 12433, Application US/10156761
; Publication No. US20030119018A1
; GENERAL INFORMATION:
; APPLICANT: OMURA, SATOSHI
; APPLICANT: IKEDA, HARUO
; APPLICANT: ISHIKAWA, JUN
; APPLICANT: HORIKAWA, HIROSHI
; APPLICANT: SHIBA, TADAYOSHI
; APPLICANT: SAKAKI, YOSHIYUKI
; APPLICANT: HATTORI, KASAHIRA
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
; FILE REFERENCE: 249-262
; CURRENT APPLICATION NUMBER: US/10/156,761
; CURRENT FILING DATE: 2002-05-29
; PRIOR APPLICATION NUMBER: JP 2001-204089
; PRIOR FILING DATE: 2001-05-30
; PRIOR APPLICATION NUMBER: JP 2001-272697
; PRIOR FILING DATE: 2001-08-02
; NUMBER OF SEQ ID NOS: 15109
; SEQ ID NO 12433
; LENGTH: 413
; TYPE: PRT

; ORGANISM: Streptomyces avermitilis
US-10-156-761-12433

Query Match      1.5%; Score 8; DB 14; Length 413;
Best Local Similarity 100.0%; Pred. No. 1e+02;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy      49 SCALRDVS 56
Db      297 SCALRDVS 304

RESULT 33
US-10-369-493-21212
; Sequence 21212, Application US/10369493
; Publication No. US20030233675A1
; GENERAL INFORMATION:
; APPLICANT: Cao, Yongwei
; APPLICANT: Hinkle, Gregory J.
; APPLICANT: Slater, Steven C.
; APPLICANT: Goldman, Barry S.
; APPLICANT: Chen, Xianfeng
; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF
; PLANTS WITH IMPROVED PROPERTIES
; FILE REFERENCE: 38-10(52052)B
; CURRENT APPLICATION NUMBER: US/10/369,493
; CURRENT FILING DATE: 2003-02-28
; PRIOR APPLICATION NUMBER: US 60/360,039
; PRIOR FILING DATE: 2002-02-21
; NUMBER OF SEQ ID NOS: 47374
; SEQ ID NO 21212
; LENGTH: 447
; TYPE: PRT
; ORGANISM: Xenorhabdus nematophilus
US-10-369-493-21212

Query Match      1.5%; Score 8; DB 15; Length 447;
Best Local Similarity 100.0%; Pred. No. 1.1e+02;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy      149 AKGLGKEI 156
Db      397 AKGLGKEI 404

RESULT 34
US-10-335-977-7865
; Sequence 7865, Application US/10335977
; Publication No. US20040052799A1
; GENERAL INFORMATION:
; APPLICANT: DOUGLAS SMITH et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES
; RELATING TO HELICOBACTER PYLORI FOR
; DIAGNOSTICS AND THERAPEUTICS
; NUMBER OF SEQUENCES: 10031
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: LAHIVE & COCKFIELD
; STREET: 28 State Street
; CITY: Boston
; STATE: Massachusetts
; COUNTRY: USA
; ZIP: 02109-1875
; COMPUTER READABLE FORM:
; MEDIUM TYPE: CD-ROM ISO9660
; COMPUTER: IBM PC Compatible
; OPERATING SYSTEM: Windows NT 4.0
; SOFTWARE: UNIX
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/10/335,977
; FILING DATE: 30-Dec-2002
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/993,002
; FILING DATE: 17-DEC-1997
```

```

; ORIGINAL SOURCE:
; ORGANISM: Helicobacter pylori
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (B) LOCATION 1...451
; SEQUENCE DESCRIPTION: SEQ ID NO: 7866:
US-10-335-977-7866

Query Match      1.5%; Score 8; DB 12; Length 451;
Best Local Similarity 100.0%; Pred. No. 1.1e+02;
Matches      8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      63 LEDILSTY 70
DB      284 LEDILSTY 291

RESULT 36
US-09-738-626-6430
; Sequence 6430, Application US/09738626
; Publication No. US20020197605A1
; GENERAL INFORMATION:
; APPLICANT: NAKAGAWA, SATOSHI
; APPLICANT: MIZOGUCHI, HIROSHI
; APPLICANT: ANDO, SEIKO
; APPLICANT: HAYASHI, MIKIRO
; APPLICANT: OCHIAI, KEIKO
; APPLICANT: YOKOI, HARUHIKO
; APPLICANT: TATEISHI, NAOKO
; APPLICANT: SENOH, AKIHIRO
; APPLICANT: IKEDA, MASATO
; APPLICANT: OZAKI, AKIO
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
; FILE REFERENCE: 249-125
; CURRENT APPLICATION NUMBER: US/09/739,626
; CURRENT FILING DATE: 2000-12-18
; PRIOR APPLICATION NUMBER: JP 99/377484
; PRIOR FILING DATE: 1999-12-16
; PRIOR APPLICATION NUMBER: JP 00/159162
; PRIOR FILING DATE: 2000-04-07
; PRIOR APPLICATION NUMBER: JP 00/280988
; PRIOR FILING DATE: 2000-08-03
; NUMBER OF SEQ ID NOS: 7059
; SOFTWARE: PatentIn ver. 3.0
; SEQ ID NO 6430
; LENGTH: 484
; TYPE: PRT
; ORGANISM: Corynebacterium glutamicum
US-09-738-626-6430

Query Match      1.5%; Score 8; DB 9; Length 484;
Best Local Similarity 100.0%; Pred. No. 1.2e+02;
Matches      8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      181 AELLSEHR 188
DB      80 AELLSEHR 87

RESULT 37
US-10-369-493-23425
; Sequence 23425, Application US/10369493
; Publication No. US20030233675A1
; GENERAL INFORMATION:
; APPLICANT: Cao, Yongwei
; APPLICANT: Hinkie, Gregory J.
; APPLICANT: Slater, Steven C.
; APPLICANT: Goldman, Barry S.
; APPLICANT: Chen, Xianfeng
; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION
; TITLE OF INVENTION: PLANTS WITH IMPROVED PROPERTIES
; FILE REFERENCE: 38-10(52052)B
; CURRENT APPLICATION NUMBER: US/10/369,493

```

CURRENT FILING DATE: 2003-02-28
PRIOR APPLICATION NUMBER: US 60/360,039
PRIOR FILING DATE: 2002-02-21
NUMBER OF SEQ ID NOS: 47374
SEQ ID NO 23425
LENGTH: 525
TYPE: PRT
ORGANISM: Deinococcus radiodurans
US-10-369-493-23425

Query Match
Best Local Similarity 1.5%; Score 8; DB 15; Length 525;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 496 APSSPRVT 503
Db 45 APSSPRVT 52

RESULT 38
US-09-962-055-41
Sequence 41, Application US/09962055
Patent No. US20020052033A1
GENERAL INFORMATION:
APPLICANT: Lees, Ann M.
Lees, Robert S.
Law, Simon W.
Arjona, Anibal A.
TITLE OF INVENTION: NOVEL LOW DENSITY LIPOPROTEIN BINDING AND TREATING ATHEROSCLEROSIS
NUMBER OF SEQUENCES: 42
CORRESPONDENCE ADDRESS:
ADDRESSEE: Fish & Richardson P.C.
STREET: 225 Franklin Street
CITY: Boston
STATE: MA
COUNTRY: USA
ZIP: 02110-2804
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: FastSEQ for Windows Version 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/962,055
FILING DATE: 24-Sep-2001
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/979,608
FILING DATE: 26-NOV-1997
APPLICATION NUMBER: US 60/031,930
FILING DATE: 27-NOV-1996
ATTORNEY/AGENT INFORMATION:
NAME: Myers, Louis
REGISTRATION NUMBER: 35,965
REFERENCE/DOCKET NUMBER: 10797-002001 (formerly 3983/59818)
TELECOMMUNICATION INFORMATION:
TELEPHONE: 617/542-5070
TELEFAX: 617/542-8906
INFORMATION FOR SEQ ID NO: 41:
SEQUENCE CHARACTERISTICS:
LENGTH: 7 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
SEQUENCE DESCRIPTION: SEQ ID NO: 41:
US-09-962-055-41

Query Match
Best Local Similarity 1.3%; Score 7; DB 9; Length 7;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 53 RDVSEEL 59
Db 1 RDVSEEL 7

RESULT 39
US-09-976-740-41
Sequence 41, Application US/09976740
Publication No. US20020194633A1
GENERAL INFORMATION:
APPLICANT: Lees, Ann M.
Lees, Robert S.
Law, Simon W.
Arjona, Anibal A.
TITLE OF INVENTION: NOVEL LOW DENSITY LIPOPROTEIN BINDING AND TREATING ATHEROSCLEROSIS
FILE REFERENCE: 10797-004001
CURRENT APPLICATION NUMBER: US/09/976,740
CURRENT FILING DATE: 2001-10-12
PRIOR APPLICATION NUMBER: 09/616,289
PRIOR FILING DATE: 2000-07-14
PRIOR APPLICATION NUMBER: US 08/979,608
PRIOR FILING DATE: 1997-11-26
PRIOR APPLICATION NUMBER: US 60/031,930
PRIOR FILING DATE: 1996-11-27
PRIOR APPLICATION NUMBER: US 60/048,547
PRIOR FILING DATE: 1997-06-03
NUMBER OF SEQ ID NOS: 53
SOFTWARE: FastSEQ for Windows Version 4.0
SEQ ID NO 41
LENGTH: 7
TYPE: PRT
ORGANISM: Homo sapiens
US-09-976-740-41

Query Match
Best Local Similarity 1.3%; Score 7; DB 9; Length 7;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 53 RDVSEEL 59
Db 1 RDVSEEL 7

RESULT 40
US-10-671-242-41
Sequence 41, Application US/10671242
Publication No. US20040040049A1
GENERAL INFORMATION:
APPLICANT: Lees, Ann M.
Lees, Robert S.
Law, Simon W.
Arjona, Anibal A.
TITLE OF INVENTION: NOVEL LOW DENSITY LIPOPROTEIN BINDING AND TREATING ATHEROSCLEROSIS
FILE REFERENCE: 10797-004001
CURRENT APPLICATION NUMBER: US/10/671,242
CURRENT FILING DATE: 2003-09-24
PRIOR APPLICATION NUMBER: US/09/616,289
PRIOR FILING DATE: 2000-07-14
PRIOR APPLICATION NUMBER: US 09/517,849
PRIOR FILING DATE: 2000-03-02
PRIOR APPLICATION NUMBER: US 08/979,608
PRIOR FILING DATE: 1997-11-26
PRIOR APPLICATION NUMBER: US 60/031,930
PRIOR FILING DATE: 1996-11-27
PRIOR APPLICATION NUMBER: US 60/048,547
PRIOR FILING DATE: 1997-06-03
NUMBER OF SEQ ID NOS: 53
SOFTWARE: FastSEQ for Windows Version 4.0
SEQ ID NO 41
LENGTH: 7

```
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-671-242-41

Query Match
Best Local Similarity 1.3%; Score 7; DB 12; Length 7;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 53 RDVSEEL 59
Db 1 RDVSEEL 7

RESULT 41
US-10-023-529-41
; Sequence 41, Application US/10023529
; Publication No. US20020129388A1
; GENERAL INFORMATION:
; APPLICANT: Lees, Robert S.
; APPLICANT: Lees, Robert S.
; APPLICANT: Law, Simon W.
; APPLICANT: Arjona, Anibal A.
; TITLE OF INVENTION: NOVEL LOW DENSITY LIPOPROTEIN BINDING
; TITLE OF INVENTION: PROTEINS AND THEIR USE IN DIAGNOSING AND TREATING
; TITLE OF INVENTION: ATHEROSCLEROSIS
; FILE REFERENCE: 10797-004001
; CURRENT APPLICATION NUMBER: US/10/023,529
; CURRENT FILING DATE: 2001-12-17
; PRIOR APPLICATION NUMBER: 09/616,289
; PRIOR FILING DATE: 2000-07-14
; PRIOR APPLICATION NUMBER: US 09/517,849
; PRIOR FILING DATE: 2000-03-02
; PRIOR APPLICATION NUMBER: US 08/979,608
; PRIOR FILING DATE: 1997-11-26
; PRIOR APPLICATION NUMBER: US 60/031,930
; PRIOR FILING DATE: 1996-11-27
; PRIOR APPLICATION NUMBER: US 60/048,547
; PRIOR FILING DATE: 1997-06-03
; NUMBER OF SEQ ID NOS: 53
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 41
; LENGTH: 7
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-023-529-41

Query Match
Best Local Similarity 1.3%; Score 7; DB 13; Length 7;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 53 RDVSEEL 59
Db 1 RDVSEEL 7

RESULT 42
US-10-023-523-41
; Sequence 41, Application US/10023523
; Publication No. US20020152485A1
; GENERAL INFORMATION:
; APPLICANT: Lees, Robert S.
; APPLICANT: Lees, Robert S.
; APPLICANT: Law, Simon W.
; APPLICANT: Arjona, Anibal A.
; TITLE OF INVENTION: NOVEL LOW DENSITY LIPOPROTEIN BINDING
; TITLE OF INVENTION: PROTEINS AND THEIR USE IN DIAGNOSING AND TREATING
; TITLE OF INVENTION: ATHEROSCLEROSIS
; FILE REFERENCE: 10797-004001
; CURRENT APPLICATION NUMBER: US/10/023,523
; CURRENT FILING DATE: 2001-12-17
; PRIOR APPLICATION NUMBER: US/09/616,289
; PRIOR FILING DATE: 2000-07-14
; PRIOR APPLICATION NUMBER: US 09/517,849
```

```
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-671-242-41

Query Match
Best Local Similarity 1.3%; Score 7; DB 12; Length 7;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 53 RDVSEEL 59
Db 1 RDVSEEL 7

RESULT 41
US-10-023-529-41
; Sequence 41, Application US/10023529
; Publication No. US20020129388A1
; GENERAL INFORMATION:
; APPLICANT: Lees, Robert S.
; APPLICANT: Lees, Robert S.
; APPLICANT: Law, Simon W.
; APPLICANT: Arjona, Anibal A.
; TITLE OF INVENTION: NOVEL LOW DENSITY LIPOPROTEIN BINDING
; TITLE OF INVENTION: PROTEINS AND THEIR USE IN DIAGNOSING AND TREATING
; TITLE OF INVENTION: ATHEROSCLEROSIS
; FILE REFERENCE: 10797-004001
; CURRENT APPLICATION NUMBER: US/10/023,529
; CURRENT FILING DATE: 2001-12-17
; PRIOR APPLICATION NUMBER: 09/616,289
; PRIOR FILING DATE: 2000-07-14
; PRIOR APPLICATION NUMBER: US 09/517,849
; PRIOR FILING DATE: 2000-03-02
; PRIOR APPLICATION NUMBER: US 08/979,608
; PRIOR FILING DATE: 1997-11-26
; PRIOR APPLICATION NUMBER: US 60/031,930
; PRIOR FILING DATE: 1996-11-27
; PRIOR APPLICATION NUMBER: US 60/048,547
; PRIOR FILING DATE: 1997-06-03
; NUMBER OF SEQ ID NOS: 53
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 41
; LENGTH: 7
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-023-529-41

Query Match
Best Local Similarity 1.3%; Score 7; DB 13; Length 7;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 53 RDVSEEL 59
Db 1 RDVSEEL 7

RESULT 42
US-10-023-523-41
; Sequence 41, Application US/10023523
; Publication No. US20020152485A1
; GENERAL INFORMATION:
; APPLICANT: Lees, Robert S.
; APPLICANT: Lees, Robert S.
; APPLICANT: Law, Simon W.
; APPLICANT: Arjona, Anibal A.
; TITLE OF INVENTION: NOVEL LOW DENSITY LIPOPROTEIN BINDING
; TITLE OF INVENTION: PROTEINS AND THEIR USE IN DIAGNOSING AND TREATING
; TITLE OF INVENTION: ATHEROSCLEROSIS
; FILE REFERENCE: 10797-004001
; CURRENT APPLICATION NUMBER: US/10/023,523
; CURRENT FILING DATE: 2001-12-17
; PRIOR APPLICATION NUMBER: US/09/616,289
; PRIOR FILING DATE: 2000-07-14
; PRIOR APPLICATION NUMBER: US 09/517,849
```

```
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-671-242-41

Query Match
Best Local Similarity 1.3%; Score 7; DB 12; Length 7;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 53 RDVSEEL 59
Db 1 RDVSEEL 7

RESULT 41
US-10-023-529-41
; Sequence 41, Application US/10616187
; Publication No. US20040013668A1
; GENERAL INFORMATION:
; APPLICANT: Lees, Robert S.
; APPLICANT: Lees, Robert S.
; APPLICANT: Law, Simon W.
; APPLICANT: Arjona, Anibal A.
; TITLE OF INVENTION: NOVEL LOW DENSITY LIPOPROTEIN BINDING
; TITLE OF INVENTION: PROTEINS AND THEIR USE IN DIAGNOSING AND TREATING
; TITLE OF INVENTION: ATHEROSCLEROSIS
; FILE REFERENCE: 10797-004001
; CURRENT APPLICATION NUMBER: US/10/616,187
; CURRENT FILING DATE: 2003-07-09
; PRIOR APPLICATION NUMBER: US/09/616,289
; PRIOR FILING DATE: 2000-07-14
; PRIOR APPLICATION NUMBER: US 09/517,849
; PRIOR FILING DATE: 2000-03-02
; PRIOR APPLICATION NUMBER: US 08/979,608
; PRIOR FILING DATE: 1997-11-26
; PRIOR APPLICATION NUMBER: US 60/031,930
; PRIOR FILING DATE: 1996-11-27
; PRIOR APPLICATION NUMBER: US 60/048,547
; NUMBER OF SEQ ID NOS: 53
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 41
; LENGTH: 7
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-616-187-41

Query Match
Best Local Similarity 1.3%; Score 7; DB 15; Length 7;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 53 RDVSEEL 59
Db 1 RDVSEEL 7

RESULT 44
US-10-225-567A-2071
; Sequence 2071, Application US/10225567A
; Publication No. US20030113798A1
; GENERAL INFORMATION:
; APPLICANT: Lifespan Biosciences
; APPLICANT: Brown, Joseph P.
```

```

; APPLICANT: Burner, Glenna C.
; APPLICANT: Roush, Christine L.
; TITLE OF INVENTION: ANTIGENIC PEPTIDES AND ANTIBODIES FOR G PROTEIN-COUPLED RECEPTORS
; FILE REFERENCE: 1920-4-4
; CURRENT APPLICATION NUMBER: US/10/225,567A
; CURRENT FILING DATE: 2001-12-19
; PRIOR APPLICATION NUMBER: 60/257,144
; PRIOR FILING DATE: 2000-12-19
; NUMBER OF SEQ ID NOS: 2292
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 2071
; LENGTH: 18
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-225-567A-2071

Query Match
Best Local Similarity 100.0%; DB 14; Length 18;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 57 BELSRQL 63
Db 6 BELSRQL 12

RESULT 45
US-10-195-730-348
; Sequence 348, Application US/10195730
; Publication No. US2003014492A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et. al
; TITLE OF INVENTION: 101 Human Secreted Proteins
; FILE REFERENCE: P2017P1
; CURRENT APPLICATION NUMBER: US/10/195,730
; CURRENT FILING DATE: 2002-07-16
; PRIOR APPLICATION NUMBER: US/09/281,976
; PRIOR FILING DATE: 1999-03-31
; PRIOR APPLICATION NUMBER: 60/060,837
; PRIOR FILING DATE: 1997-10-02
; PRIOR APPLICATION NUMBER: 60/060,852
; PRIOR FILING DATE: 1997-10-02
; NUMBER OF SEQ ID NOS: 390
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 348
; LENGTH: 27
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-195-730-348

Query Match
Best Local Similarity 100.0%; DB 81; Length 27;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 427 LLEMAEE 433
Db 5 LLEMAEE 11

RESULT 46
US-09-864-761-36055
; Sequence 36055, Application US/09864761
; Patent No. US20020048763A1
; GENERAL INFORMATION:
; APPLICANT: Penn, Sharon G.
; APPLICANT: Rank, David R.
; APPLICANT: Hanzel, David K.
; APPLICANT: Chen, Wensheng
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
; FILE REFERENCE: Aecmca-X-1
; CURRENT APPLICATION NUMBER: US/09/864,761
; CURRENT FILING DATE: 2001-05-23
; PRIOR APPLICATION NUMBER: US 60/180,312
; APPLICANT: Lightowler, Joanne

; APPLICANT: Burner, Glenna C.
; APPLICANT: Roush, Christine L.
; TITLE OF INVENTION: ANTIGENIC PEPTIDES AND ANTIBODIES FOR G PROTEIN-COUPLED RECEPTORS
; FILE REFERENCE: 1920-4-4
; CURRENT APPLICATION NUMBER: US/10/225,567A
; CURRENT FILING DATE: 2001-12-19
; PRIOR APPLICATION NUMBER: 60/257,144
; PRIOR FILING DATE: 2000-12-19
; NUMBER OF SEQ ID NOS: 2292
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 2071
; LENGTH: 18
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-225-567A-2071

; PRIOR FILING DATE: 2000-02-04
; PRIOR APPLICATION NUMBER: US 60/207,456
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: US 09/532,366
; PRIOR FILING DATE: 2000-08-03
; PRIOR APPLICATION NUMBER: GB 24263.6
; PRIOR FILING DATE: 2000-10-04
; PRIOR APPLICATION NUMBER: US 60/236,359
; PRIOR FILING DATE: 2000-09-27
; PRIOR APPLICATION NUMBER: PCT/US01/00666
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00667
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00664
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00669
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00665
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00668
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00663
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00662
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00661
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00670
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: US 60/234,687
; PRIOR FILING DATE: 2000-09-21
; PRIOR APPLICATION NUMBER: US 09/608,408
; PRIOR FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: US 09/774,203
; PRIOR FILING DATE: 2001-01-29
; NUMBER OF SEQ ID NOS: 49117
; SOFTWARE: Annomax Sequence Listing Engine vers. 1.1
; SEQ ID NO 36055
; LENGTH: 38
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: MAP TO 299716.4
; OTHER INFORMATION: EXPRESSED IN BT474, SIGNAL = 2.8
; OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 1.1
; OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 1.2
; OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 1.2
; OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 0.95
; OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 0.96
; OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 2.8
; OTHER INFORMATION: EXPRESSED IN HBL100, SIGNAL = 2
; OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 2.4
; OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 1.5
; OTHER INFORMATION: EST HUMAN HIT: BE312629.1, EVALUATE 6.00e-15
; OTHER INFORMATION: SWISSPROT HIT: Q60429, EVALUATE 7.00e-16
; US-09-864-761-36055

Query Match
Best Local Similarity 100.0%; DB 9; Length 38;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 281 KLRQENM 287
Db 23 KLRQENM 29

RESULT 47
US-09-491-614-31
; Sequence 31, Application US/09491614
; Publication No. US20030100027A1
; GENERAL INFORMATION:
; APPLICANT: Coyler, John
; APPLICANT: Lightowler, Joanne
```

```
; TITLE OF INVENTION: Methods and Compositions Using Coiled Binding Partners
; FILE REFERENCE: 4256/78978
; CURRENT APPLICATION NUMBER: US/09/491,614
; CURRENT FILING DATE: 2000-01-26
; PRIOR APPLICATION NUMBER: 09/259,474
; PRIOR FILING DATE: 1999-02-26
; NUMBER OF SEQ ID NOS: 32
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 31
; LENGTH: 44
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: Peptide
; OTHER INFORMATION: construct for coiled coil derived from the folate
; OTHER INFORMATION: receptor and the consensus for GPI anchor
; OTHER INFORMATION: addition.
US-09-491-614-31

Query Match      1.3%; Score 7; DB 10; Length 44;
Best Local Similarity 100.0%; Pred. No. 1.2e+02; Indels 0; Gaps 0;
Matches 7; Conservative 0; Mismatches 0;

QY      406 IKKLEKE 412
Db      23 IKKLEKE 29

RESULT 48
US-09-986-480-335
; Sequence 335, Application US/09986480
; Publication No. US20030027999A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: 143 Human Secreted Proteins
; FILE REFERENCE: PS500P1
; CURRENT APPLICATION NUMBER: US/09/986,480
; CURRENT FILING DATE: 2001-11-08
; PRIOR APPLICATION NUMBER: PCT/US00/12788
; PRIOR FILING DATE: 2000-05-11
; PRIOR APPLICATION NUMBER: US 60/134,068
; PRIOR FILING DATE: 1999-05-13
; NUMBER OF SEQ ID NOS: 456
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 335
; LENGTH: 45
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-986-480-335

Query Match      1.3%; Score 7; DB 10; Length 45;
Best Local Similarity 100.0%; Pred. No. 1.3e+02; Indels 0; Gaps 0;
Matches 7; Conservative 0; Mismatches 0;

QY      12 EGAQERP 18
Db      34 EGAQERP 40

RESULT 49
US-09-892-877-355
; Sequence 355, Application US/09892877
; Publication No. US20030077809A1
; GENERAL INFORMATION:
; APPLICANT: Ruben et al.
; TITLE OF INVENTION: 97 Human secreted proteins
; FILE REFERENCE: P2028P1
; CURRENT APPLICATION NUMBER: US/09/892,877
; CURRENT FILING DATE: 2001-06-28
; PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: US/09/437,658
; PRIOR FILING DATE: EARLIER FILING DATE: 1999-11-10
; NUMBER OF SEQ ID NOS: 461
; SOFTWARE: PatentIn Ver. 2.0
```

```
; SEQ ID NO 355
; LENGTH: 45
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-892-877-355

Query Match      1.3%; Score 7; DB 10; Length 45;
Best Local Similarity 100.0%; Pred. No. 1.3e+02; Indels 0; Gaps 0;
Matches 7; Conservative 0; Mismatches 0;

QY      12 EGAQERP 18
Db      34 EGAQERP 40

RESULT 50
US-09-948-783-359
; Sequence 359, Application US/09948783
; Publication No. US20030100051A1
; GENERAL INFORMATION:
; APPLICANT: Ruben et al.
; TITLE OF INVENTION: 97 Human secreted proteins
; FILE REFERENCE: P2028P2
; CURRENT APPLICATION NUMBER: US/09/948,783
; CURRENT FILING DATE: 2001-09-10
; PRIOR APPLICATION NUMBER: 60/231,846
; PRIOR FILING DATE: 2000-09-11
; PRIOR APPLICATION NUMBER: 09/892,877
; PRIOR FILING DATE: 2001-06-28
; PRIOR APPLICATION NUMBER: 09/437,658
; PRIOR FILING DATE: 1999-11-10
; PRIOR APPLICATION NUMBER: PCT/US99/09847
; PRIOR FILING DATE: 1999-05-06
; PRIOR APPLICATION NUMBER: 60/085,093
; PRIOR FILING DATE: 1998-05-12
; PRIOR APPLICATION NUMBER: 60/085,094
; PRIOR FILING DATE: 1998-05-12
; PRIOR APPLICATION NUMBER: 60/085,105
; PRIOR FILING DATE: 1998-05-12
; PRIOR APPLICATION NUMBER: 60/085,180
; PRIOR FILING DATE: 1998-05-12
; PRIOR APPLICATION NUMBER: 60/085,927
; PRIOR FILING DATE: 1998-05-18
; PRIOR APPLICATION NUMBER: 60/085,906
; PRIOR FILING DATE: 1998-05-18
; PRIOR APPLICATION NUMBER: 60/085,924
; PRIOR FILING DATE: 1998-05-18
; PRIOR APPLICATION NUMBER: 60/085,922
; PRIOR FILING DATE: 1998-05-18
; PRIOR APPLICATION NUMBER: 60/085,921
; PRIOR FILING DATE: 1998-05-18
; PRIOR APPLICATION NUMBER: 60/085,923
; PRIOR FILING DATE: 1998-05-18
; PRIOR APPLICATION NUMBER: 60/085,925
; PRIOR FILING DATE: 1998-05-18
; PRIOR APPLICATION NUMBER: 60/085,928
; PRIOR FILING DATE: 1998-05-18
; PRIOR APPLICATION NUMBER: 60/085,920
; PRIOR FILING DATE: 1998-05-18
; NUMBER OF SEQ ID NOS: 465
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 359
; LENGTH: 45
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-948-783-359

Query Match      1.3%; Score 7; DB 10; Length 45;
Best Local Similarity 100.0%; Pred. No. 1.3e+02; Indels 0; Gaps 0;
Matches 7; Conservative 0; Mismatches 0;

QY      12 EGAQERP 18
Db      34 EGAQERP 40
```

Db 34 EGAQRP 40

RESULT 51
US-10-424-599-154654
; Sequence 154654, Application US/10424599
; Publication No. US20040031072A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa Thomas J
; APPLICANT: Kovalic David K
; APPLICANT: Zhou Yihua
; APPLICANT: Cao Yongwei
; TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated With
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
; FILE REFERENCE: 38-21(53223)B
; CURRENT APPLICATION NUMBER: US/10/424,599
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 285684
; SEQ ID NO 154654
; LENGTH: 55
; TYPE: PRT
; ORGANISM: Glycine max
; FEATURE:
; NAME/KEY: unsure
; LOCATION: (1)..(55)
; OTHER INFORMATION: unsure at all xaa locations
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT3847_110674C.1.pap
US-10-424-599-154654

Query Match 1.3%; Score 7; DB 12; Length 55;
Best Local Similarity 100.0%; Pred. No. 1.5e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 23 PAVEAEG 29
DB 37 PAVEAEG 43

RESULT 52
US-10-424-599-246225
; Sequence 246225, Application US/10424599
; Publication No. US20040031072A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa Thomas J
; APPLICANT: Kovalic David K
; APPLICANT: Zhou Yihua
; APPLICANT: Cao Yongwei
; TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated With
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
; FILE REFERENCE: 38-21(53223)B
; CURRENT APPLICATION NUMBER: US/10/424,599
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 285684
; SEQ ID NO 246225
; LENGTH: 78
; TYPE: PRT
; ORGANISM: Glycine max
; FEATURE:
; NAME/KEY: unsure
; LOCATION: (1)..(78)
; OTHER INFORMATION: unsure at all xaa locations
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT3847_64373C.1.pap
US-10-424-599-246225

Query Match 1.3%; Score 7; DB 12; Length 78;
Best Local Similarity 100.0%; Pred. No. 2.1e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 315 DLQQQLV 321
DB 16 DLQQQLV 22

RESULT 53
US-10-425-114-68147
; Sequence 68147, Application US/10425114
; Publication No. US20040034888A1
; GENERAL INFORMATION:
; APPLICANT: Liu, Jirsdong
; APPLICANT: Zhou, Yihua
; APPLICANT: Kovalic, David K.
; APPLICANT: Screen, Steven E
; APPLICANT: Tabaska, Jack E
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
; FILE REFERENCE: 38-21(5313)B
; CURRENT APPLICATION NUMBER: US/10/425,114
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 73128
; SEQ ID NO 68147
; LENGTH: 80
; TYPE: PRT
; ORGANISM: Zea mays
; FEATURE:
; OTHER INFORMATION: Clone ID: 701174728_FLI.pap
US-10-425-114-68147

Query Match 1.3%; Score 7; DB 12; Length 80;
Best Local Similarity 100.0%; Pred. No. 2.1e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 518 GQTGPQE 524
DB 58 GQTGPQE 64

RESULT 54
US-09-796-692-1230
; Sequence 1230, Application US/09796692
; Publication No. US20020198362A1
; GENERAL INFORMATION:
; APPLICANT: Gaiger, Alexander
; APPLICANT: Algate, Paul A.
; APPLICANT: Mannion, Jane
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE DETECTION, DIAGNOSIS AND THERAPY
; TITLE OF INVENTION: HEMATOLOGICAL MALIGNANCIES
; FILE REFERENCE: 2077.001200
; CURRENT APPLICATION NUMBER: US/09/796,692
; CURRENT FILING DATE: 2001-03-01
; PRIOR APPLICATION NUMBER: 60/186,126
; PRIOR FILING DATE: 2000-03-01
; PRIOR APPLICATION NUMBER: 60/190,479
; PRIOR FILING DATE: 2000-03-17
; PRIOR APPLICATION NUMBER: 60/200,545
; PRIOR FILING DATE: 2000-04-27
; PRIOR APPLICATION NUMBER: 60/200,303
; PRIOR FILING DATE: 2000-04-28
; PRIOR APPLICATION NUMBER: 60/200,779
; PRIOR FILING DATE: 2000-04-28
; PRIOR APPLICATION NUMBER: 60/200,999
; PRIOR FILING DATE: 2000-05-01
; PRIOR APPLICATION NUMBER: 60/202,084
; PRIOR FILING DATE: 2000-05-04
; PRIOR APPLICATION NUMBER: 60/206,201
; PRIOR FILING DATE: 2000-05-22
; PRIOR APPLICATION NUMBER: 60/218,950
; PRIOR FILING DATE: 2000-07-14
; PRIOR APPLICATION NUMBER: 60/222,903
; PRIOR FILING DATE: 2000-08-03
; PRIOR APPLICATION NUMBER: 60/223,416
; PRIOR FILING DATE: 2000-08-04
; PRIOR APPLICATION NUMBER: 60/223,378
; PRIOR FILING DATE: 2000-08-07

; NUMBER OF SEQ ID NOS: 9597
; SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 1230
; LENGTH: 82
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-796-892-1230

Query Match 1.3%; Score 7; DB 9; Length 82;
Best Local Similarity 100.0%; Pred. No. 2.2e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 399 MERKTKK 405
Db 20 MERKTKK 26

RESULT 55
US-10-040-862-1230
; Sequence 1230, Application US/10040862
; Publication No. US20030078396A1
; GENERAL INFORMATION:
; APPLICANT: Gaiger, Alexander
; APPLICANT: Mannion, Jane
; APPLICANT: Mannion, Jane
; APPLICANT: Retter, Marc
; APPLICANT: Corixa Corporation
; TITLE OF INVENTION: Compositions and Methods for the Detection, Diagnosis and Therapy
; TITLE OF INVENTION: Hematological Malignancies
; FILE REFERENCE: 014058-013520US
; CURRENT APPLICATION NUMBER: US/10/040,862
; CURRENT FILING DATE: 2001-11-06
; PRIOR APPLICATION NUMBER: US 60/186,126
; PRIOR FILING DATE: 2000-03-01
; PRIOR APPLICATION NUMBER: US 60/190,479
; PRIOR FILING DATE: 2000-03-17
; PRIOR APPLICATION NUMBER: US 60/200,545
; PRIOR FILING DATE: 2000-04-27
; PRIOR APPLICATION NUMBER: US 60/200,303
; PRIOR FILING DATE: 2000-04-28
; PRIOR APPLICATION NUMBER: US 60/200,779
; PRIOR FILING DATE: 2000-04-28
; PRIOR APPLICATION NUMBER: US 60/202,084
; PRIOR FILING DATE: 2000-05-01
; PRIOR APPLICATION NUMBER: US 60/206,201
; PRIOR FILING DATE: 2000-05-01
; PRIOR APPLICATION NUMBER: US 60/218,950
; PRIOR FILING DATE: 2000-07-14
; PRIOR APPLICATION NUMBER: US 60/222,903
; PRIOR FILING DATE: 2000-08-03
; PRIOR APPLICATION NUMBER: US 60/223,416
; PRIOR FILING DATE: 2000-08-04
; PRIOR APPLICATION NUMBER: US 60/223,378
; PRIOR FILING DATE: 2000-08-07
; PRIOR APPLICATION NUMBER: US 09/796,692
; PRIOR FILING DATE: 2001-03-01
; NUMBER OF SEQ ID NOS: 10467
; SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 1230
; LENGTH: 82
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-040-862-1230

Query Match 1.3%; Score 7; DB 14; Length 82;
Best Local Similarity 100.0%; Pred. No. 2.2e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 399 MERKTKK 405
Db 20 MERKTKK 26

RESULT 56
US-10-057-475B-1230
; Sequence 1230, Application US/10057475B
; Publication No. US20040002068A1
; GENERAL INFORMATION:
; APPLICANT: Gaiger, Alexander
; APPLICANT: Mannion, Jane
; APPLICANT: Mannion, Jane
; APPLICANT: Clapper, Jonathan David
; APPLICANT: Wang, Aijun
; APPLICANT: Ordonez, Nadia
; APPLICANT: Carter, Lauren
; APPLICANT: McNeill, Patricia Dianne
; APPLICANT: Corixa Corporation
; TITLE OF INVENTION: Compositions and Methods for the Detection, Diagnosis and Therapy
; TITLE OF INVENTION: Hematological Malignancies
; FILE REFERENCE: 014058-014402US
; CURRENT APPLICATION NUMBER: US/10/057,475B
; CURRENT FILING DATE: 2002-01-22
; PRIOR APPLICATION NUMBER: US 60/186,126
; PRIOR FILING DATE: 2000-03-01
; PRIOR APPLICATION NUMBER: US 60/190,479
; PRIOR FILING DATE: 2000-03-17
; PRIOR APPLICATION NUMBER: US 60/200,545
; PRIOR FILING DATE: 2000-04-27
; PRIOR APPLICATION NUMBER: US 60/200,303
; PRIOR FILING DATE: 2000-04-28
; PRIOR APPLICATION NUMBER: US 60/200,779
; PRIOR FILING DATE: 2000-04-28
; PRIOR APPLICATION NUMBER: US 60/200,999
; PRIOR FILING DATE: 2000-05-01
; PRIOR APPLICATION NUMBER: US 60/202,084
; PRIOR FILING DATE: 2000-05-04
; PRIOR APPLICATION NUMBER: US 60/206,201
; PRIOR FILING DATE: 2000-05-22
; PRIOR APPLICATION NUMBER: US 60/218,950
; PRIOR FILING DATE: 2000-07-14
; PRIOR APPLICATION NUMBER: US 60/222,903
; PRIOR FILING DATE: 2000-08-03
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 10979
; SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 1230
; LENGTH: 82
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-057-475B-1230

Query Match 1.3%; Score 7; DB 15; Length 82;
Best Local Similarity 100.0%; Pred. No. 2.2e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 399 MERKTKK 405
Db 20 MERKTKK 26

RESULT 57
US-10-154-884B-1230
; Sequence 1230, Application US/10154884B
; Publication No. US20040005561A1
; GENERAL INFORMATION:
; APPLICANT: Gaiger, Alexander
; APPLICANT: Mannion, Jane
; APPLICANT: Mannion, Jane
; APPLICANT: Retter, Marc W.
; APPLICANT: Corixa Corporation
; TITLE OF INVENTION: Compositions and Methods for the Detection, Diagnosis and Therapy
; TITLE OF INVENTION: Hematological Malignancies
; FILE REFERENCE: 014058-013521US
; CURRENT APPLICATION NUMBER: US/10/154,884B

; CURRENT FILING DATE: 2002-05-23
; PRIOR APPLICATION NUMBER: US 60/186,126
; PRIOR FILING DATE: 2000-03-01
; PRIOR APPLICATION NUMBER: US 60/190,479
; PRIOR FILING DATE: 2000-03-17
; PRIOR APPLICATION NUMBER: US 60/200,545
; PRIOR FILING DATE: 2000-04-27
; PRIOR APPLICATION NUMBER: US 60/200,303
; PRIOR FILING DATE: 2000-04-28
; PRIOR APPLICATION NUMBER: US 60/200,779
; PRIOR FILING DATE: 2000-04-28
; PRIOR APPLICATION NUMBER: US 60/200,999
; PRIOR FILING DATE: 2000-05-01
; PRIOR APPLICATION NUMBER: US 60/202,084
; PRIOR FILING DATE: 2000-05-04
; PRIOR APPLICATION NUMBER: US 60/206,201
; PRIOR FILING DATE: 2000-05-22
; PRIOR APPLICATION NUMBER: US 60/218,950
; PRIOR FILING DATE: 2000-07-14
; PRIOR APPLICATION NUMBER: US 60/222,903
; PRIOR FILING DATE: 2000-08-03
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 11290
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 1230
; LENGTH: 82
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-154-884B-1230

Query Match 1.3%; Score 7; DB 15; Length 82;
Best Local Similarity 100.0%; Pred. No. 2.2e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 399 MEKMTKK 405
Db 20 MEKMTKK 26

RESULT 58
US-10-424-599-197528
; Sequence 197528, Application US/10424599
; Publication No. US20040031072A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa Thomas J
; APPLICANT: Kovalic David K
; APPLICANT: Zhou Yitua
; APPLICANT: Cao Yongwei
; TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated With
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
; FILE REFERENCE: 38-21(53223)B
; CURRENT APPLICATION NUMBER: US/10/424,599
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 285684
; SEQ ID NO 197528
; LENGTH: 89
; TYPE: PRT
; ORGANISM: Glycine max
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT3847_20394C.1.pep
US-10-424-599-197528

Query Match 1.3%; Score 7; DB 12; Length 89;
Best Local Similarity 100.0%; Pred. No. 2.3e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 219 VLARSKL 225
Db 21 VLARSKL 27

RESULT 59
US-09-796-692-1236

; Sequence 1236, Application US/09796692
; Publication No. US20020198362A1
; GENERAL INFORMATION:
; APPLICANT: Gaiger, Alexander
; APPLICANT: Algate, Paul A.
; APPLICANT: Mannion, Jane
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE DETECTION, DIAGNOSIS AND THERAP
; TITLE OF INVENTION: HEMATOLOGICAL MALIGNANCIES
; FILE REFERENCE: 2077.001200
; CURRENT APPLICATION NUMBER: US/09/796,692
; CURRENT FILING DATE: 2001-03-01
; PRIOR APPLICATION NUMBER: 60/186,126
; PRIOR FILING DATE: 2000-03-01
; PRIOR APPLICATION NUMBER: 60/190,479
; PRIOR FILING DATE: 2000-03-17
; PRIOR APPLICATION NUMBER: 60/200,545
; PRIOR FILING DATE: 2000-04-27
; PRIOR APPLICATION NUMBER: 60/200,303
; PRIOR FILING DATE: 2000-04-28
; PRIOR APPLICATION NUMBER: 60/200,779
; PRIOR FILING DATE: 2000-04-28
; PRIOR APPLICATION NUMBER: 60/200,999
; PRIOR FILING DATE: 2000-05-01
; PRIOR APPLICATION NUMBER: 60/202,084
; PRIOR FILING DATE: 2000-05-04
; PRIOR APPLICATION NUMBER: 60/206,201
; PRIOR FILING DATE: 2000-05-22
; PRIOR APPLICATION NUMBER: 60/218,950
; PRIOR FILING DATE: 2000-07-14
; PRIOR APPLICATION NUMBER: 60/222,903
; PRIOR FILING DATE: 2000-08-03
; PRIOR APPLICATION NUMBER: 60/223,416
; PRIOR FILING DATE: 2000-08-04
; PRIOR APPLICATION NUMBER: 60/223,378
; PRIOR FILING DATE: 2000-08-07
; NUMBER OF SEQ ID NOS: 9597
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 1236
; LENGTH: 91
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-796-692-1236

Query Match 1.3%; Score 7; DB 9; Length 91;
Best Local Similarity 100.0%; Pred. No. 2.4e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 399 MEKMTKK 405
Db 20 MEKMTKK 26

RESULT 60
US-09-796-692-1693
; Sequence 1693, Application US/09796692
; Publication No. US20020198362A1
; GENERAL INFORMATION:
; APPLICANT: Gaiger, Alexander
; APPLICANT: Algate, Paul A.
; APPLICANT: Mannion, Jane
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE DETECTION, DIAGNOSIS AND THERAP
; TITLE OF INVENTION: HEMATOLOGICAL MALIGNANCIES
; FILE REFERENCE: 2077.001200
; CURRENT APPLICATION NUMBER: US/09/796,692
; CURRENT FILING DATE: 2001-03-01
; PRIOR APPLICATION NUMBER: 60/186,126
; PRIOR FILING DATE: 2000-03-01
; PRIOR APPLICATION NUMBER: 60/190,479
; PRIOR FILING DATE: 2000-03-17
; PRIOR APPLICATION NUMBER: 60/200,545
; PRIOR FILING DATE: 2000-04-27
; PRIOR APPLICATION NUMBER: 60/200,303
; PRIOR FILING DATE: 2000-04-28

;; PRIOR APPLICATION NUMBER: US 60/206,201
;; PRIOR FILING DATE: 2000-05-22
;; PRIOR APPLICATION NUMBER: US 60/218,950
;; PRIOR FILING DATE: 2000-07-14
;; PRIOR APPLICATION NUMBER: US 60/222,903
;; PRIOR FILING DATE: 2000-08-03
;; Remaining Prior Application data removed - See File Wrapper or PALM.
;; NUMBER OF SEQ ID NOS: 11290
;; SOFTWARE: FastSeq for Windows Version 3.0
;; SEQ ID NO 1236
;; LENGTH: 91
;; TYPE: PRT
;; ORGANISM: Homo sapiens
US-10-154-884B-1236

Query Match 1.3%; Score 7; DB 15; Length 91;
Best Local Similarity 100.0%; Pred. No. 2.4e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 399 MERMTKK 405

DB 20 MERMTKK 26

RESULT 66

US-10-154-884B-1693
;; Sequence 1693, Application US/10154884B
;; Publication No. US20040005561A1
;; GENERAL INFORMATION:
;; APPLICANT: Gaiger, Alexander
;; APPLICANT: Algate, Paul A.
;; APPLICANT: Mannion, Jane
;; APPLICANT: Retter, Marc W.
;; APPLICANT: Corixa Corporation
;; TITLE OF INVENTION: Compositions and Methods for the Detection, Diagnosis and Therapy
;; FILE REFERENCE: 014058-013521US
;; CURRENT APPLICATION NUMBER: US/10/154,884B
;; CURRENT FILING DATE: 2002-05-23
;; PRIOR APPLICATION NUMBER: US 60/186,126
;; PRIOR FILING DATE: 2000-03-01
;; PRIOR APPLICATION NUMBER: US 60/190,479
;; PRIOR FILING DATE: 2000-03-17
;; PRIOR APPLICATION NUMBER: US 60/200,545
;; PRIOR FILING DATE: 2000-04-27
;; PRIOR APPLICATION NUMBER: US 60/200,303
;; PRIOR FILING DATE: 2000-04-28
;; PRIOR APPLICATION NUMBER: US 60/200,779
;; PRIOR FILING DATE: 2000-04-28
;; PRIOR APPLICATION NUMBER: US 60/200,999
;; PRIOR FILING DATE: 2000-05-01
;; PRIOR APPLICATION NUMBER: US 60/202,084
;; PRIOR FILING DATE: 2000-05-04
;; PRIOR APPLICATION NUMBER: US 60/206,201
;; PRIOR FILING DATE: 2000-05-22
;; PRIOR APPLICATION NUMBER: US 60/218,950
;; PRIOR FILING DATE: 2000-07-14
;; PRIOR APPLICATION NUMBER: US 60/222,903
;; PRIOR FILING DATE: 2000-08-03
;; Remaining Prior Application data removed - See File Wrapper or PALM.
;; NUMBER OF SEQ ID NOS: 11290
;; SOFTWARE: FastSeq for Windows Version 3.0
;; SEQ ID NO 1693
;; LENGTH: 91
;; TYPE: PRT
;; ORGANISM: Homo sapiens
US-10-154-884B-1693

Query Match 1.3%; Score 7; DB 15; Length 91;
Best Local Similarity 100.0%; Pred. No. 2.4e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 399 MERMTKK 405

DB 20 MERMTKK 26

RESULT 67

US-09-864-761-39633
;; Sequence 39633, Application US/09864761
;; Patent No. US20020048763A1
;; GENERAL INFORMATION:
;; APPLICANT: Penn, Sharron G.
;; APPLICANT: Rank, David R.
;; APPLICANT: Hanzel, David K.
;; APPLICANT: Chen, Wensheng
;; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
;; FILE REFERENCE: Aemica-X-1
;; CURRENT APPLICATION NUMBER: US/09/864,761
;; CURRENT FILING DATE: 2001-05-23
;; PRIOR APPLICATION NUMBER: US 60/180,312
;; PRIOR FILING DATE: 2000-02-04
;; PRIOR APPLICATION NUMBER: US 60/207,456
;; PRIOR FILING DATE: 2000-05-26
;; PRIOR APPLICATION NUMBER: US 09/632,366
;; PRIOR FILING DATE: 2000-08-03
;; PRIOR APPLICATION NUMBER: GB 24263.6
;; PRIOR FILING DATE: 2000-10-04
;; PRIOR APPLICATION NUMBER: US 60/236,359
;; PRIOR FILING DATE: 2000-09-27
;; PRIOR APPLICATION NUMBER: PCT/US01/00666
;; PRIOR FILING DATE: 2001-01-30
;; PRIOR APPLICATION NUMBER: PCT/US01/00667
;; PRIOR FILING DATE: 2001-01-30
;; PRIOR APPLICATION NUMBER: PCT/US01/00664
;; PRIOR FILING DATE: 2001-01-30
;; PRIOR APPLICATION NUMBER: PCT/US01/00669
;; PRIOR FILING DATE: 2001-01-30
;; PRIOR APPLICATION NUMBER: PCT/US01/00665
;; PRIOR FILING DATE: 2001-01-30
;; PRIOR APPLICATION NUMBER: PCT/US01/00668
;; PRIOR FILING DATE: 2001-01-30
;; PRIOR APPLICATION NUMBER: PCT/US01/00663
;; PRIOR FILING DATE: 2001-01-30
;; PRIOR APPLICATION NUMBER: PCT/US01/00662
;; PRIOR FILING DATE: 2001-01-30
;; PRIOR APPLICATION NUMBER: PCT/US01/00661
;; PRIOR FILING DATE: 2001-01-30
;; PRIOR APPLICATION NUMBER: PCT/US01/00670
;; PRIOR FILING DATE: 2001-01-30
;; PRIOR APPLICATION NUMBER: US 60/234,687
;; PRIOR FILING DATE: 2000-09-21
;; PRIOR APPLICATION NUMBER: US 09/608,408
;; PRIOR FILING DATE: 2000-06-30
;; PRIOR APPLICATION NUMBER: US 09/774,203
;; PRIOR FILING DATE: 2001-01-29
;; NUMBER OF SEQ ID NOS: 49117
;; SOFTWARE: Anomax Sequence Listing Engine vers. 1.1
;; SEQ ID NO 39633
;; LENGTH: 95
;; TYPE: PRT
;; ORGANISM: Homo sapiens
;; FEATURE:
;; OTHER INFORMATION: MAP TO AC002465.1
;; OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 2.9
;; OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 0.69
;; OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 0.42
US-09-864-761-39633

Query Match 1.3%; Score 7; DB 9; Length 95;
Best Local Similarity 100.0%; Pred. No. 2.5e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 249 EEEERKK 255

```
Db 58 EBEKEK 64

RESULT 68
US-09-804-014A-56
; Sequence 56, Application US/09804014A
; Publication No. US20030064489A1
; GENERAL INFORMATION:
; APPLICANT: Li, Li
; APPLICANT: Padigaru, Muralidhara
; APPLICANT: Vernet, Corine
; APPLICANT: Fernandes, Elma
; APPLICANT: Shimkets, Richard
; APPLICANT: Spaderna, Steven
; APPLICANT: Majumder, Kumud
; TITLE OF INVENTION: Novel Polypeptides and Nucleic Acids Encoding Same
; FILE REFERENCES: 15966-721 US
; CURRENT APPLICATION NUMBER: US/09/804,014A
; CURRENT FILING DATE: 2002-04-24
; PRIOR APPLICATION NUMBER: 60/188,316
; PRIOR FILING DATE: 2000-03-10
; PRIOR APPLICATION NUMBER: 60/188,277
; PRIOR FILING DATE: 2000-03-10
; PRIOR APPLICATION NUMBER: 60/189,139
; PRIOR FILING DATE: 2000-03-14
; PRIOR APPLICATION NUMBER: 60/189,140
; PRIOR FILING DATE: 2000-03-14
; PRIOR APPLICATION NUMBER: 60/190,401
; PRIOR FILING DATE: 2000-03-17
; PRIOR APPLICATION NUMBER: 60/190,231
; PRIOR FILING DATE: 2000-03-17
; NUMBER OF SEQ ID NOS: 75
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 56
; LENGTH: 99
; TYPE: PRT
; ORGANISM: Rattus norvegicus
US-09-804-014A-56

Query Match 1.3%; Score 7; DB 12; Length 99;
Best Local Similarity 100.0%; Pred. No. 2.6e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 241 EBGVQRA 247
Db 2 EBGVQRA 8

RESULT 69
US-10-424-599-266203
; Sequence 266203, Application US/10424599
; Publication No. US20040031072A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa Thomas J
; APPLICANT: Kovalic David K
; APPLICANT: Zhou Yihua
; APPLICANT: Cao Yongwei
; TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated With
; FILE REFERENCES: 38-21(53223)B
; CURRENT APPLICATION NUMBER: US/10/424,599
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 285684
; SEQ ID NO 266203
; LENGTH: 100
; TYPE: PRT
; ORGANISM: Glycine max
; FEATURE:
; NAME/KEY: unsure
; LOCATION: (1)..(100)
; OTHER INFORMATION: unsure at all xaa locations
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT3847_82402C.1.pep

Db 58 EBEKEK 64

RESULT 68
US-09-804-014A-56
; Sequence 56, Application US/09804014A
; Publication No. US20030064489A1
; GENERAL INFORMATION:
; APPLICANT: Li, Li
; APPLICANT: Padigaru, Muralidhara
; APPLICANT: Vernet, Corine
; APPLICANT: Fernandes, Elma
; APPLICANT: Shimkets, Richard
; APPLICANT: Spaderna, Steven
; APPLICANT: Majumder, Kumud
; TITLE OF INVENTION: Novel Polypeptides and Nucleic Acids Encoding Same
; FILE REFERENCES: 15966-721 US
; CURRENT APPLICATION NUMBER: US/09/804,014A
; CURRENT FILING DATE: 2002-04-24
; PRIOR APPLICATION NUMBER: 60/188,316
; PRIOR FILING DATE: 2000-03-10
; PRIOR APPLICATION NUMBER: 60/188,277
; PRIOR FILING DATE: 2000-03-10
; PRIOR APPLICATION NUMBER: 60/189,139
; PRIOR FILING DATE: 2000-03-14
; PRIOR APPLICATION NUMBER: 60/189,140
; PRIOR FILING DATE: 2000-03-14
; PRIOR APPLICATION NUMBER: 60/190,401
; PRIOR FILING DATE: 2000-03-17
; PRIOR APPLICATION NUMBER: 60/190,231
; PRIOR FILING DATE: 2000-03-17
; NUMBER OF SEQ ID NOS: 75
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 56
; LENGTH: 99
; TYPE: PRT
; ORGANISM: Rattus norvegicus
US-09-804-014A-56

Query Match 1.3%; Score 7; DB 12; Length 99;
Best Local Similarity 100.0%; Pred. No. 2.6e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 241 EBGVQRA 247
Db 2 EBGVQRA 8

RESULT 69
US-10-424-599-266203
; Sequence 266203, Application US/10424599
; Publication No. US20040031072A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa Thomas J
; APPLICANT: Kovalic David K
; APPLICANT: Zhou Yihua
; APPLICANT: Cao Yongwei
; TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated With
; FILE REFERENCES: 38-21(53223)B
; CURRENT APPLICATION NUMBER: US/10/424,599
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 285684
; SEQ ID NO 266203
; LENGTH: 100
; TYPE: PRT
; ORGANISM: Glycine max
; FEATURE:
; NAME/KEY: unsure
; LOCATION: (1)..(100)
; OTHER INFORMATION: unsure at all xaa locations
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT3847_82402C.1.pep
```

```
US-10-424-599-266203

Query Match 1.3%; Score 7; DB 12; Length 100;
Best Local Similarity 100.0%; Pred. No. 2.6e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 291 ERLKELI 297
Db 45 ERLKELI 51

RESULT 70
US-10-424-599-153810
; Sequence 153810, Application US/10424599
; Publication No. US20040031072A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa Thomas J
; APPLICANT: Kovalic David K
; APPLICANT: Zhou Yihua
; APPLICANT: Cao Yongwei
; TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated With
; FILE REFERENCES: 38-21(53223)B
; CURRENT APPLICATION NUMBER: US/10/424,599
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 285684
; SEQ ID NO 153810
; LENGTH: 101
; TYPE: PRT
; ORGANISM: Glycine max
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT3847_109913C.1.pep
US-10-424-599-153810

Query Match 1.3%; Score 7; DB 12; Length 101;
Best Local Similarity 100.0%; Pred. No. 2.6e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 246 RAREEEE 252
Db 55 RAREEEE 61

RESULT 71
US-10-264-049-4293
; Sequence 4293, Application US/10264049
; Publication No. US20040005579A1
; GENERAL INFORMATION:
; APPLICANT: Birse et al.
; FILE REFERENCES: PAL33P1
; CURRENT APPLICATION NUMBER: US/10/264,049
; CURRENT FILING DATE: 2002-10-04
; PRIOR APPLICATION NUMBER: PCI/US01/18569
; PRIOR FILING DATE: 2001-06-07
; PRIOR APPLICATION NUMBER: US 60/209,467
; PRIOR FILING DATE: 2000-06-07
; NUMBER OF SEQ ID NOS: 4369
; SOFTWARE: PatentIn Ver. 3.1
; SEQ ID NO 4293
; LENGTH: 106
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: MISC_FEATURE
; LOCATION: (34)
; OTHER INFORMATION: Xaa equals any of the twenty naturally occurring L-amino acids
; FEATURE:
; NAME/KEY: MISC_FEATURE
; LOCATION: (63)
; OTHER INFORMATION: Xaa equals any of the twenty naturally occurring L-amino acids
; FEATURE:
; NAME/KEY: MISC_FEATURE
```

; LOCATION: (106)
; OTHER INFORMATION: Xaa equals any of the twenty naturally occurring L-amino acids
US-10-264-049-4293

Query Match 1.3%; Score 7; DB 15; Length 106;
Best Local Similarity 100.0%; Pred. No. 2.7e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 477 QGSLTDS 483
Db 81 QGSLTDS 87

RESULT 72

US-10-195-730-345
; Sequence 345, Application US/10195730
; Publication No. US20030144492A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et. al
; TITLE OF INVENTION: 101 Human Secreted Proteins
; FILE REFERENCE: P201721
; CURRENT APPLICATION NUMBER: US/10/195,730
; CURRENT FILING DATE: 2002-07-16
; PRIOR APPLICATION NUMBER: US/09/281,976
; PRIOR FILING DATE: 1999-03-31
; PRIOR APPLICATION NUMBER: 60/060,837
; PRIOR FILING DATE: 1997-10-02
; PRIOR APPLICATION NUMBER: 60/060,862
; PRIOR FILING DATE: 1997-10-02
; NUMBER OF SEQ ID NOS: 390
; SOFTWARE: Patent in Ver. 2.0
; SEQ ID NO 345
; LENGTH: 110
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-195-730-345

Query Match 1.3%; Score 7; DB 14; Length 110;
Best Local Similarity 100.0%; Pred. No. 2.8e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 427 LLEMAEE 433
Db 81 LLEMAEE 87

RESULT 73

US-10-425-114-50441
; Sequence 50441, Application US/10425114
; Publication No. US20040034888A1
; GENERAL INFORMATION:
; APPLICANT: Liu, Jingdong
; APPLICANT: Zhou, Yihua
; APPLICANT: Kovalic, David K.
; APPLICANT: Screen, Steven E
; APPLICANT: Tabaska, Jack E
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
; FILE REFERENCE: 38-21(53313)B
; CURRENT APPLICATION NUMBER: US/10/425,114
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 73128
; SEQ ID NO 50441
; LENGTH: 113
; TYPE: PRT
; ORGANISM: Zea mays
; FEATURES:
; OTHER INFORMATION: Clone ID: LIB3079-022-B2_FLI.pep
US-10-425-114-50441

Query Match 1.3%; Score 7; DB 12; Length 113;
Best Local Similarity 100.0%; Pred. No. 2.9e+02;

Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 151 GLGKEIT 157
Db 20 GLGKEIT 26

RESULT 74

US-10-424-599-165582
; Sequence 165582, Application US/10424599
; Publication No. US20040031072A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa Thomas J
; APPLICANT: Kovalic David K
; APPLICANT: Zhou Yihua
; APPLICANT: Cao Yongwei
; TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated With
; FILE REFERENCE: 38-21(53223)B
; CURRENT APPLICATION NUMBER: US/10/424,599
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 285684
; SEQ ID NO 165582
; LENGTH: 114
; TYPE: PRT
; ORGANISM: Glycine max
; FEATURES:
; OTHER INFORMATION: Clone ID: PAT_MRT3847_120536C.1.pap
US-10-424-599-165582

Query Match 1.3%; Score 7; DB 12; Length 114;
Best Local Similarity 100.0%; Pred. No. 2.9e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 236 NRSLEKE 242
Db 45 NRSLEKE 51

RESULT 75

US-10-424-599-196307
; Sequence 196307, Application US/10424599
; Publication No. US20040031072A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa Thomas J
; APPLICANT: Kovalic David K
; APPLICANT: Zhou Yihua
; APPLICANT: Cao Yongwei
; TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated With
; FILE REFERENCE: 38-21(53223)B
; CURRENT APPLICATION NUMBER: US/10/424,599
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 285684
; SEQ ID NO 196307
; LENGTH: 115
; TYPE: PRT
; ORGANISM: Glycine max
; FEATURES:
; NAME/KEY: unsure
; LOCATION: (1)_(115)
; OTHER INFORMATION: unsure at all Xaa locations
; FEATURES:
; OTHER INFORMATION: Clone ID: PAT_MRT3847_19282C.1.pap
US-10-424-599-196307

Query Match 1.3%; Score 7; DB 12; Length 115;
Best Local Similarity 100.0%; Pred. No. 2.9e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 250 EBEKKE 256
Db 97 EBEKKE 103

```

;
; MOLECULE TYPE: protein
; HYPOTHETICAL: YES
; ORIGINAL SOURCE:
; ORGANISM: Helicobacter pylori
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (B) LOCATION 1...117
; SEQUENCE DESCRIPTION: SEQ ID NO: 4955:
US-10-335-977-4955

Query Match      1.3%; Score 7; DB 12; Length 117;
Best Local Similarity 100.0%; Pred. No. 3e+02; Mismatches 0; Indels 0; Gaps 0;
Matches 7; Conservative 0;

QY      406 IKKLEKE 412
DB      4 IKKLEKE 10
      |||||

RESULT 78
US-10-424-599-241987
; Sequence 241987, Application US/10424599
; Publication No. US20040031072A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa Thomas J
; APPLICANT: Kovalic David K
; APPLICANT: Zhou Yihua
; APPLICANT: Cao Yongwei
; TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated With
; FILE OF INVENTION: Plants and Uses Thereof for Plant Improvement
; FILE REFERENCE: 38-21(53223)B
; CURRENT APPLICATION NUMBER: US/10/424,599
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 285684
; SEQ ID NO 241987
; LENGTH: 119
; TYPE: PRT
; ORGANISM: Glycine max
; FEATURE:
; NAME/KEY: unsure
; LOCATION: (1)...(119)
; OTHER INFORMATION: unsure at all Xaa locations
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT3847_60541C.1.pep
US-10-424-599-241987

Query Match      1.3%; Score 7; DB 12; Length 119;
Best Local Similarity 100.0%; Pred. No. 3e+02; Mismatches 0; Indels 0; Gaps 0;
Matches 7; Conservative 0;

QY      127 EEIRQSD 133
DB      93 EEIRQSD 99
      |||||

RESULT 79
US-10-443-622-121
; Sequence 121, Application US/10443622
; Publication No. US20040024192A1
; GENERAL INFORMATION:
; APPLICANT: Carter et al.
; TITLE OF INVENTION: 19 Human Secreted Proteins
; FILE REFERENCE: P2009P1
; CURRENT APPLICATION NUMBER: US/10/443,622
; CURRENT FILING DATE: 2003-05-23
; PRIOR APPLICATION NUMBER: PCT/US98/13608
; PRIOR FILING DATE: 1998-06-30
; PRIOR APPLICATION NUMBER: 60/051,480
; PRIOR FILING DATE: 1997-07-01
; PRIOR APPLICATION NUMBER: 60/051,381
; PRIOR FILING DATE: 1997-07-01
; PRIOR APPLICATION NUMBER: 60/058,663
; PRIOR FILING DATE: 1997-09-12

```

```

;
; MOLECULE TYPE: protein
; HYPOTHETICAL: YES
; ORIGINAL SOURCE:
; ORGANISM: Helicobacter pylori
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (B) LOCATION 1...117
; SEQUENCE DESCRIPTION: SEQ ID NO: 4955:
US-10-335-977-4955

Query Match      1.3%; Score 7; DB 14; Length 115;
Best Local Similarity 100.0%; Pred. No. 2.9e+02; Mismatches 0; Indels 0; Gaps 0;
Matches 7; Conservative 0;

QY      314 KDLQOOL 320
DB      86 KDLQOOL 92
      |||||

RESULT 77
US-10-335-977-4955
; Sequence 4955, Application US/10335977
; Publication No. US20040052799A1
; GENERAL INFORMATION:
; APPLICANT: DOUGLAS SMITH et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES
; FILE OF INVENTION: RELATING TO HELICOBACTER PYLORI FOR
; FILE REFERENCE: 38-21(53223)B
; CURRENT APPLICATION NUMBER: US/10/424,599
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 10031
; CORRESPONDENCE ADDRESS:
; ADDRESS: LAHIVE & COCKFIELD
; STREET: 28 State Street
; CITY: Boston
; STATE: Massachusetts
; COUNTRY: USA
; ZIP: 02109-1875
; COMPUTER READABLE FORM:
; MEDIUM TYPE: CD-ROM ISO9660
; COMPUTER: IBM PC Compatible
; OPERATING SYSTEM: Windows NT 4.0
; SOFTWARE: UNIX
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/10/335,977
; FILING DATE: 30-Dec-2002
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/993,002
; FILING DATE: 17-DEC-1997
; ATTORNEY/AGENT INFORMATION:
; NAME: Mandragoras, Amy E.
; REGISTRATION NUMBER: 36,207
; REFERENCE/DOCKET NUMBER: GTN-018
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617) 227-7400
; TELEFAX: (617) 742-4214
; INFORMATION FOR SEQ ID NO: 4955:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 117 amino acids
; TYPE: amino acid
; TOPOLOGY: linear

```

```
; PRIOR APPLICATION NUMBER: 60/058,598
; PRIOR FILING DATE: 1997-09-12
; NUMBER OF SEQ ID NOS: 156
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 121
; LENGTH: 122
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-443-622-121

Query Match          1.3%; Score 7; DB 16; Length 122;
Best Local Similarity 100.0%; Pred. No. 3.1e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 468 RVQDLISA 474
Db 4 RVQDLISA 10

RESULT 80
US-10-424-599-150351
; Sequence 150351, Application US/10424599
; Publication No. US20040031072A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa Thomas J
; APPLICANT: Kowalic David K
; APPLICANT: Zhou Yihua
; APPLICANT: Cao Yongwei
; TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated With
; FILE REFERENCE: 38-21(53223)B
; CURRENT APPLICATION NUMBER: US/10/424,599
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 285684
; SEQ ID NO 150351
; LENGTH: 124
; TYPE: PRT
; ORGANISM: Glycine max
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT3847_10678C.1.pep
US-10-424-599-150351

Query Match          1.3%; Score 7; DB 12; Length 124;
Best Local Similarity 100.0%; Pred. No. 3.1e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 403 TKIKKL 409
Db 17 TKIKKL 23

RESULT 81
US-08-849-303-19
; Sequence 19, Application US/08849303
; Publication No. US20030221209A1
; GENERAL INFORMATION:
; APPLICANT: Atkinson, Howard J.
; APPLICANT: McPherson, Michael J.
; APPLICANT: Urwin, Peter E.
; TITLE OF INVENTION: MODIFIED PROTEINASE INHIBITORS
; NUMBER OF SEQUENCES: 79
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Klauber & Jackson
; STREET: 411 Hackensack Avenue, 4th Floor
; CITY: Hackensack
; STATE: New Jersey
; COUNTRY: USA
; ZIP: 07601
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
```

```
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/849,303
; FILING DATE: 21-MAY-1997
; CLASSIFICATION: 514
; ATTORNEY/AGENT INFORMATION:
; NAME: Jackson Esq., David A.
; REGISTRATION NUMBER: 26,742
; REFERENCE/DOCKET NUMBER: 1321-1-003
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 201-487-5800
; TELEFAX: 201-343-1684
; TELEX: 133521
; INFORMATION FOR SEQ ID NO: 19:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 127 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; HYPOTHETICAL: NO
US-08-849-303-19

Query Match          1.3%; Score 7; DB 8; Length 127;
Best Local Similarity 100.0%; Pred. No. 3.2e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 241 EBGVQRA 247
Db 27 EBGVQRA 33

RESULT 82
US-10-108-260A-3437
; Sequence 3437, Application US/10108260A
; Publication No. US20040005560A1
; GENERAL INFORMATION:
; APPLICANT: HELIX RESEARCH INSTITUTE
; TITLE OF INVENTION: NO. US20040005560A1 full length cDNA
; FILE REFERENCE: HI-A0106
; CURRENT APPLICATION NUMBER: US/10/108,260A
; CURRENT FILING DATE: 2002-03-27
; NUMBER OF SEQ ID NOS: 5458
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 3437
; LENGTH: 130
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-108-260A-3437

Query Match          1.3%; Score 7; DB 15; Length 130;
Best Local Similarity 100.0%; Pred. No. 3.3e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 488 RPEGPGA 494
Db 77 RPEGPGA 83

RESULT 83
US-10-335-977-4956
; Sequence 4956, Application US/10335977
; Publication No. US20040052799A1
; GENERAL INFORMATION:
; APPLICANT: DOUGLAS SMITH et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES
; RELATING TO HELICOBACTER PYLORI FOR
; DIAGNOSTICS AND THERAPEUTICS
; NUMBER OF SEQUENCES: 10031
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: LAHIVE & COCKFIELD
; STREET: 28 State Street
; CITY: Boston
; STATE: Massachusetts
```



```
; ORGANISM: Helicobacter pylori
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (B) LOCATION 1...137
; SEQUENCE DESCRIPTION: SEQ ID NO: 4958:
US-10-335-977-4958

Query Match      1.3%; Score 7; DB 12; Length 137;
Best Local Similarity 100.0%; Pred. No. 3.4e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 406 IKKLEKE 412
DB 8 IKKLEKE 14

RESULT 86
US-09-847-208-130
; Sequence 130, Application US/09847208
; Publication No. US20030082190A1
; GENERAL INFORMATION:
; APPLICANT: Saxon, Andrew
; APPLICANT: Zhang, Ke
; APPLICANT: Zhu, Daocheng
; TITLE OF INVENTION: FUSION MOLECULES AND TREATMENT OF
; TITLE OF INVENTION: IGE-MEDIATED ALLERGIC DISEASES
; FILE REFERENCE: UC67.002A
; CURRENT APPLICATION NUMBER: US/09/847,208
; CURRENT FILING DATE: 2001-05-01
; NUMBER OF SEQ ID NOS: 177
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 130
; LENGTH: 138
; TYPE: PRT
; ORGANISM: Parietaria judaica
US-09-847-208-130

Query Match      1.3%; Score 7; DB 10; Length 138;
Best Local Similarity 100.0%; Pred. No. 3.4e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 116 EKEPSKG 122
DB 58 EKEPSKG 64

RESULT 87
US-09-847-208-128
; Sequence 128, Application US/09847208
; Publication No. US20030082190A1
; GENERAL INFORMATION:
; APPLICANT: Saxon, Andrew
; APPLICANT: Zhang, Ke
; APPLICANT: Zhu, Daocheng
; TITLE OF INVENTION: FUSION MOLECULES AND TREATMENT OF
; TITLE OF INVENTION: IGE-MEDIATED ALLERGIC DISEASES
; FILE REFERENCE: UC67.002A
; CURRENT APPLICATION NUMBER: US/09/847,208
; CURRENT FILING DATE: 2001-05-01
; NUMBER OF SEQ ID NOS: 177
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 128
; LENGTH: 139
; TYPE: PRT
; ORGANISM: Parietaria judaica
US-09-847-208-128

Query Match      1.3%; Score 7; DB 10; Length 139;
Best Local Similarity 100.0%; Pred. No. 3.5e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 116 EKEPSKG 122
DB 58 EKEPSKG 64

RESULT 88
US-10-425-114-69680
; Sequence 69680, Application US/10425114
; Publication No. US2004003488A1
; GENERAL INFORMATION:
; APPLICANT: Liu, Jingdong
; APPLICANT: Zhou, Yihua
; APPLICANT: Kovalic, David K.
; APPLICANT: Screen, Steven B.
; APPLICANT: Tabaska, Jack E
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
; FILE REFERENCE: 38-21(53313)B
; CURRENT APPLICATION NUMBER: US/10/425,114
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 73128
; SEQ ID NO 69680
; LENGTH: 142
; TYPE: PRT
; ORGANISM: Zea mays
; FEATURE:
; OTHER INFORMATION: Clone ID: UC-ZMFLMO17176H10_FLI.pep
US-10-425-114-69680

Query Match      1.3%; Score 7; DB 12; Length 142;
Best Local Similarity 100.0%; Pred. No. 3.5e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 202 SOLVQEK 208
DB 105 SOLVQEK 111

RESULT 89
US-10-424-599-167688
; Sequence 167688, Application US/10424599
; Publication No. US20040031072A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa Thomas J
; APPLICANT: Kovalic David K
; APPLICANT: Zhou Yihua
; APPLICANT: Cao Yongwei
; TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated With
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
; FILE REFERENCE: 38-21(53223)B
; CURRENT APPLICATION NUMBER: US/10/424,599
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 285684
; SEQ ID NO 167688
; LENGTH: 146
; TYPE: PRT
; ORGANISM: Glycine max
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT3847_122439C.1.pep
US-10-424-599-167688

Query Match      1.3%; Score 7; DB 12; Length 146;
Best Local Similarity 100.0%; Pred. No. 3.6e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 423 SNKALLE 429
DB 59 SNKALLE 65

RESULT 90
US-09-893-519A-11
; Sequence 11, Application US/09893519A
; Publication No. US20030027243A1
```

GENERAL INFORMATION:
; APPLICANT: ANADYS PHARMACEUTICALS, INC.
; APPLICANT: THOMPSON, Craig
; APPLICANT: MOORE, Jeffrey
; APPLICANT: BUURMAN, Ed T.
; APPLICANT: BRADLEY, John
; APPLICANT: DESILVA, Thamara
; APPLICANT: HARRIS, Sandra
; APPLICANT: KOMARNITSKY, Svetlana
; APPLICANT: MENDILLO, Marc
; APPLICANT: MOORE, Daniel
; APPLICANT: MCCOY, Melissa
; APPLICANT: SANDERSON, Karen
; APPLICANT: HAO, Tariq
; APPLICANT: ZHU, Shuhao
; APPLICANT: LONG, Fan
; APPLICANT: DAVIDOV, Eugene
; TITLE OF INVENTION: ANTIFUNGAL COMPOUNDS AND METHODS OF USE
; FILE REFERENCE: 0342/1G548-US2
; CURRENT APPLICATION NUMBER: US/09/893,519A
; CURRENT FILING DATE: 2001-06-28
; PRIOR APPLICATION NUMBER: US 60/215,164
; PRIOR FILING DATE: 2000-06-29
; PRIOR APPLICATION NUMBER: US 60/224,457
; PRIOR FILING DATE: 2000-08-10
; NUMBER OF SEQ ID NOS: 146
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 11
; LENGTH: 156
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc feature
; OTHER INFORMATION: Corresponds to SEQ ID NO: 84
; PUBLICATION INFORMATION:
; DATABASE ACCESSION NUMBER: Human Genbank/AA042873
; DATABASE ENTRY DATE: 2000-01-05
; RELEVANT RESIDUES: (1)...(156)
US-09-893-519A-11

Query Match 1.3%; Score 7; DB 10; Length 156;
Best Local Similarity 100.0%; Pred. No. 3.8e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 59 LSRQLED 65
DB 7 LSRQLED 13

RESULT 91
US-09-934-289A-20
; Sequence 20, Application US/09934289A
; Patent No. US20020132297A1
; GENERAL INFORMATION:
; APPLICANT: Busfield, Samantha J.
; TITLE OF INVENTION: NOVEL MOLECULES OF THE
; TITLE OF INVENTION: HERPESVIRUS-ENTRY-MEDIATOR-RELATED
; TITLE OF INVENTION: PROTEIN FAMILY AND USES THEREOF
; FILE REFERENCE: MEIO98-061C1P1(N)
; CURRENT APPLICATION NUMBER: US/09/934,289A
; CURRENT FILING DATE: 2001-08-21
; PRIOR APPLICATION NUMBER: US 09/342,767
; PRIOR FILING DATE: 1999-06-29
; PRIOR APPLICATION NUMBER: US 09/146,950
; PRIOR FILING DATE: 1998-09-03
; NUMBER OF SEQ ID NOS: 58
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 20
; LENGTH: 159
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-934-289A-20

Query Match 1.3%; Score 7; DB 9; Length 159;
Best Local Similarity 100.0%; Pred. No. 3.9e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 145 EKKKAKG 151
DB 153 EKKKAKG 159

RESULT 92
US-10-424-599-163595
; Sequence 163595, Application US/10424599
; Publication No. US20040031072A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa Thomas J
; APPLICANT: Kovalic David K
; APPLICANT: Zhou Yihua
; APPLICANT: Cao Yongwei
; TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated With
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
; FILE REFERENCE: 38-21(53223)B
; CURRENT APPLICATION NUMBER: US/10/424,599
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 285684
; SEQ ID NO 163595
; LENGTH: 166
; TYPE: PRT
; ORGANISM: Glycine max
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT3847_118745C.1.pep
US-10-424-599-163595

Query Match 1.3%; Score 7; DB 12; Length 166;
Best Local Similarity 100.0%; Pred. No. 4.1e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 137 DRDHRP 143
DB 71 DRDHRP 77

RESULT 93
US-10-424-599-241739
; Sequence 241739, Application US/10424599
; Publication No. US20040031072A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa Thomas J
; APPLICANT: Kovalic David K
; APPLICANT: Zhou Yihua
; APPLICANT: Cao Yongwei
; TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated With
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
; FILE REFERENCE: 38-21(53223)B
; CURRENT APPLICATION NUMBER: US/10/424,599
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 285684
; SEQ ID NO 241739
; LENGTH: 167
; TYPE: PRT
; ORGANISM: Glycine max
; FEATURE:
; NAME/KEY: unsure
; LOCATION: (1)...(167)
; OTHER INFORMATION: unsure at all Xaa locations
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT3847_60317C.1.pep
US-10-424-599-241739

Query Match 1.3%; Score 7; DB 12; Length 167;
Best Local Similarity 100.0%; Pred. No. 4.1e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 143 PQEKKKA 149

```
Db      85 PQKCKKA 91
|||||
RESULT 94
US-10-424-599-262163
; Sequence 262163, Application US/10424599
; Publication No. US20040031072A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa Thomas J
; APPLICANT: Kovalic David K
; APPLICANT: Zhou Yihua
; APPLICANT: Cao Yongwei
; TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated with
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
; FILE REFERENCE: 38-21(53223)B
; CURRENT APPLICATION NUMBER: US/10/424,599
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 285684
; SEQ ID NO 262163
; LENGTH: 171
; TYPE: PRT
; ORGANISM: Glycine max
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT3847_78755C.1.pep
US-10-424-599-262163

Query Match      1.3%; Score 7; DB 12; Length 171;
Best Local Similarity 100.0%; Pred. No. 4.2e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy      90 EPEDAEX 96
|||||
Db      130 EPEDAEX 136

RESULT 95
US-09-847-208-129
; Sequence 129, Application US/09847208
; Publication No. US20030082190A1
; GENERAL INFORMATION:
; APPLICANT: Saxon, Andrew
; APPLICANT: Zhang, Ke
; APPLICANT: Zhu, Daocheng
; TITLE OF INVENTION: FUSION MOLECULES AND TREATMENT OF
; TITLE OF INVENTION: IGE-MEDIATED ALLERGIC DISEASES
; FILE REFERENCE: UC67.002A
; CURRENT APPLICATION NUMBER: US/09/847,208
; CURRENT FILING DATE: 2001-05-01
; NUMBER OF SEQ ID NOS: 177
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 129
; LENGTH: 176
; TYPE: PRT
; ORGANISM: Parietaria judaica
US-09-847-208-129

Query Match      1.3%; Score 7; DB 10; Length 176;
Best Local Similarity 100.0%; Pred. No. 4.3e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy      116 EKEPSKG 122
|||||
Db      59 EKEPSKG 65

RESULT 96
US-10-276-774-2467
; Sequence 2467, Application US/10276774
; Publication No. US2004005345A1
; GENERAL INFORMATION:
; APPLICANT: Hyseq, Inc.
; APPLICANT: Tang, Y, Tom et al
```

```
; TITLE OF INVENTION: No. US20040053245A1el Nucleic Acids and Polypeptides
; FILE REFERENCE: 21272-030
; CURRENT APPLICATION NUMBER: US/10/276,774
; CURRENT FILING DATE: 2002-11-18
; PRIOR APPLICATION NUMBER: 09/560,875
; PRIOR FILING DATE: 2000-04-27
; PRIOR APPLICATION NUMBER: 09/496,914
; PRIOR FILING DATE: 2000-02-03
; NUMBER OF SEQ ID NOS: 2700
; SOFTWARE: Custom
; SEQ ID NO 2467
; LENGTH: 186
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (1)...(186)
; OTHER INFORMATION: Xaa = any amino acid or nothing
US-10-276-774-2467

Query Match      1.3%; Score 7; DB 12; Length 186;
Best Local Similarity 100.0%; Pred. No. 4.5e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy      477 QGSLTDS 483
|||||
Db      47 QGSLTDS 53

RESULT 97
US-09-934-289A-18
; Sequence 18, Application US/09934289A
; Patent No. US20020132297A1
; GENERAL INFORMATION:
; APPLICANT: Busfield, Samantha J.
; TITLE OF INVENTION: NOVEL MOLECULES OF THE
; TITLE OF INVENTION: HERPESVIRUS-ENTRY-MEDIATOR-RELATED
; FILE REFERENCE: PB1098-061CPCNI (M)
; CURRENT APPLICATION NUMBER: US/09/934,289A
; CURRENT FILING DATE: 2001-08-21
; PRIOR APPLICATION NUMBER: US 09/342,767
; PRIOR FILING DATE: 1999-06-29
; PRIOR APPLICATION NUMBER: US 09/146,950
; PRIOR FILING DATE: 1998-09-03
; NUMBER OF SEQ ID NOS: 58
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 18
; LENGTH: 197
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: SIGNAL
; LOCATION: (1)...(38)
US-09-934-289A-18

Query Match      1.3%; Score 7; DB 9; Length 197;
Best Local Similarity 100.0%; Pred. No. 4.7e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy      145 EKKVAKG 151
|||||
Db      191 EKKVAKG 197

RESULT 98
US-10-425-114-45419
; Sequence 45419, Application US/10425114
; Publication No. US2004003488A1
; GENERAL INFORMATION:
; APPLICANT: Liu, Jingdong
; APPLICANT: Zhou, Yihua
; APPLICANT: Kovalic, David K.
```

```
; APPLICANT: Screen, Steven E
; APPLICANT: Tabaska, Jack E
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
; FILE REFERENCE: 38-21(53313)B
; CURRENT APPLICATION NUMBER: US/10/425,114
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 73128
; SEQ ID NO 45419
; LENGTH: 206
; TYPE: PRT
; ORGANISM: Zea mays
; FEATURE:
; OTHER INFORMATION: Clone ID: LIB3088-030-A1_FLI.pep
US-10-425-114-45419

Query Match      1.3%; Score 7; DB 12; Length 206;
Best Local Similarity 100.0%; Pred.No. 4.9e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 396 KOEMERK 402
Db 115 KOEMERK 121

RESULT 99
US-10-104-047-2607
; Sequence 2607, Application US/10104047
; Publication No. US20030236392A1
; GENERAL INFORMATION:
; APPLICANT: HELIX RESEARCH INSTITUTE
; FILE REFERENCE: H1-A0105
; CURRENT APPLICATION NUMBER: US/10/104,047
; CURRENT FILING DATE: 2002-03-25
; PRIOR APPLICATION NUMBER:
; PRIOR FILING DATE:
; NUMBER OF SEQ ID NOS: 4096
; SOFTWARE: Patentin Ver. 2.1
; SEQ ID NO 2607
; LENGTH: 207
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-104-047-2607

Query Match      1.3%; Score 7; DB 15; Length 207;
Best Local Similarity 100.0%; Pred.No. 4.9e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 148 KAKGLGK 154
Db 126 KAKGLGK 132

RESULT 100
US-10-424-599-244841
; Sequence 244841, Application US/10424599
; Publication No. US20040031072A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa Thomas J
; APPLICANT: Kovalic David K
; APPLICANT: Zhou Yihua
; APPLICANT: Cao Yongwei
; TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated With
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
; FILE REFERENCE: 38-21(53223)B
; CURRENT APPLICATION NUMBER: US/10/424,599
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 285684
; SEQ ID NO 244841
; LENGTH: 208
; TYPE: PRT
```

```
; ORGANISM: Glycine max
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT3847_63120C.1.pep
US-10-424-599-244841

Query Match      1.3%; Score 7; DB 12; Length 208;
Best Local Similarity 100.0%; Pred.No. 5e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 246 RAREEES 252
Db 44 RAREEES 50

Search completed: June 7, 2004, 14:48:11
Job time : 46 secs
```